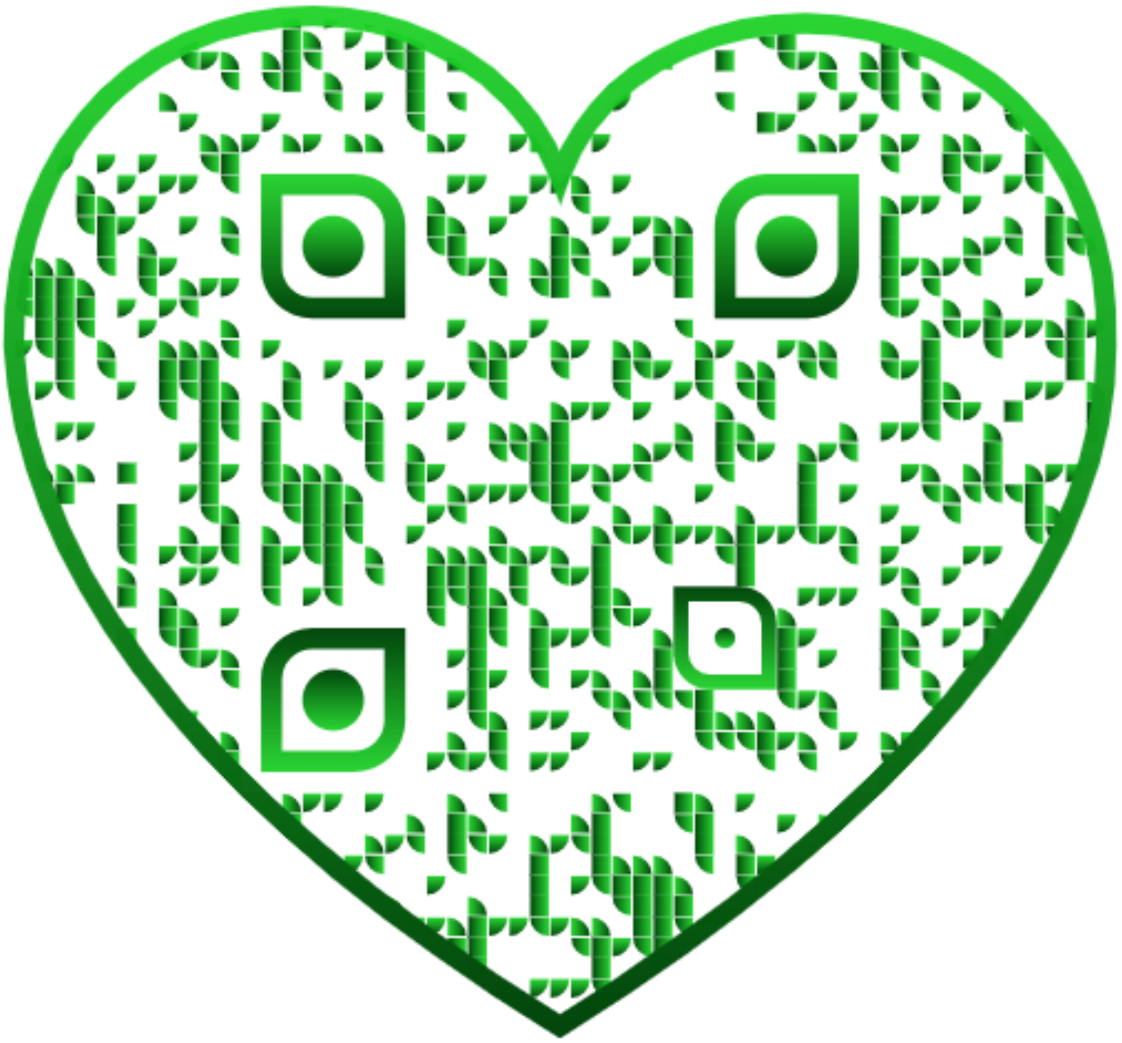
Antoine Denoix

BIG DATA, SMART DATA, STUPID DATA...

Comment (vraiment) valoriser vos données

DUNOD



Du même auteur :

L'Affiliation, Dunod, 2010.

Webanalyse, Dunod, 2013.

Illustrations : Jérôme Denoix

Pictogrammes : Arthur Shlain/ Noun Project

Eliricon/Noun Project

Akin Demir/Noun Project

Ngamlerdelk.design/Noun Project

Couverture : Atelier Didier Thimonier

Maquette : PCA

Le pictogramme qui figure ci-contre mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du photocopillage.
Le Code de la propriété intellectuelle du 1er juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée dans les établissements d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée.
Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation de l'auteur, de son éditeur ou du Centre français d'exploitation du droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris).

© Dunod, 2018

11 rue Paul Bert, 92240 Malakoff
www.dunod.com
ISBN 978-2-10-077746-4

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2° et 3° a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).
Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

Copyright © 2018 Dunod.

À France et Emmanuelle,
arrivées parmi nous un 14 juillet 2017,
au moment de la rédaction de l'ouvrage.
Que la vie, pour vous, soit belle
comme une mélodie de Jean Ferrat !

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

Copyright © 2018 Dunod.

Sommaire

PRÉFACE XI
INTRODUCTION XIII

CHAPITRE 1 DÉFINIR VOTRE STRATÉGIE 1

1 La data, pour mieux faire ! 3
2 La data, pour faire autrement ! 13
3 La data, pour ne pas disparaître ! 23

CHAPITRE 2 BÂTIR UNE CULTURE DE LA DATA 31

1 La data, ce n'est pas une arme ! 33
2 La collecte pour tous ! 41
3 La data, un patrimoine vivant ! 48

CHAPITRE 3 CONCEVOIR VOTRE TRAJECTOIRE 57

1 Le Grand Soir technologique, c'est ce soir ! 61
2 Voir grand, très grand ! 69
3 Humain, trop humain ! 74

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

CHAPITRE 4 LIBÉRER VOS ÉNERGIES 79

1 Bousculer, sans braquer ! 81
2 La diversité, avant tout ! 85
3 Un mandat clair, sinon rien ! 93

CHAPITRE 5 RÉUSSIR VOTRE EXÉCUTION 97

1 Le dernier mètre, tu réussiras ! 99
2 De la réglementation, tu te préoccuperas ! 103
3 Les algorithmes, tu ne sophistiqueras pas ! 106
4 Un projet informatico-informatique, tu éviteras ! 109
5 Sur les data externes, tu ne fantasmeras pas ! 111
6 Les data digitales, tu chériras ! 115
7 Tes clients, tu associeras ! 118
8 Décrire et prédire, tu ne confondras pas ! 123
9 De la data, tu te méfieras ! 129

CONCLUSION 135
BIBLIOGRAPHIE 137
WEBOGRAPHIE 139

Copyright © 2018 Dunod.

Demain, tout, absolument tout, produira de la data. Ceux qui sauront s'en servir réussiront. Les autres disparaîtront. La data, ce n'est pas un nouveau *buzzword*, c'est une lame de fond, aussi puissante que le digital. Le problème ? Les gains *business* tardent souvent à venir. Recruter quelques *Data Scientists* pour développer un algorithme et attendre la fumée blanche au-dessus de la salle de réunion, c'est l'échec assuré. Une stratégie data, c'est un voyage collectif. Pour réussir, c'est toute votre organisation qu'il faut bousculer : vos procédures, vos talents, votre culture… jusqu'à votre proposition de valeur ! Faites donc fi des algorithmes trop sophistiqués. C'est dans la capacité de vos équipes opérationnelles à s'approprier la data que résidera votre succès. Vous souhaitez décoller, accélérer ? Stratégies, exécution, casting, contraintes réglementaires… ce petit manuel traitera de tout, sans tabou, pour vous permettre d'aller droit au but : booster la valeur *business* de vos data !

© Dunod - Toute reproduction non autorisée est un délit.
Copyright © 2018 Dunod

Copyright © 2018 Dunod.

Préface

Régulièrement, de nombreuses actualités font état de nouvelles technologies autour de la data, de nouvelles aventures entrepreneuriales et du poids croissant des GAFA dans l'économie comme dans la société. Cette nouvelle donnée, la data, est appelée à révolutionner notre économie, nos entreprises et nos modes de vie. Un véritable nouveau secteur se crée, mobilisant de nouveaux talents et expertises dont les codes peuvent bouleverser les idées établies.

Chacun d'entre nous, décideur ou acteur dans son entreprise, nous posons de nombreuses questions. Comment la data peut-elle concerner mon activité, mon entreprise ou mon équipe ? Comment créer la dynamique data en entreprise ? Comment créer une architecture et une culture data ambitieuse et pragmatique ? Comment éviter le big data gadget ?

Si nous sommes convaincus que nous sommes tous concernés, que notre entreprise pourra en tirer un avantage différenciant (surtout si nous ne perdons pas de temps), nous ne savons cependant pas comment commencer, nous nous interrogeons sur tout. Sur quel secteur faire porter prioritairement son attention et ses investissements ? Quels objectifs devons-nous nous fixer ?

Pour ceux qui ont commencé, les premiers résultats concrets sont rarement immédiats, et le management des attentes devient indispensable pour conserver l'engagement des équipes.

Copyright © 2018 Dunod
© Dunod - Toute reproduction non autorisée est un délit.

Ces questions, ces challenges et ces décisions, j'ai été amené à les rencontrer. À la tête de mes équipes, j'ai entamé ce voyage depuis quatre ans aux côtés d'Antoine. J'ai commis des erreurs, ajusté l'action et modifié les objectifs et attentes.

En lisant ce livre, je reconnais toute l'expérience, la maîtrise et l'expertise d'un homme qui a su mobiliser les équipes pour entreprendre le voyage, modifier régulièrement la route pour l'optimiser et nous permettre d'enregistrer nos premières victoires.

Beaucoup parlent de la data comme du nouvel or noir du 21e siècle. Beaucoup en parlent, mais peu s'en enrichissent. Pourquoi ? Parce que trop peu savent l'exploiter !

Avec humilité, pédagogie, finesse et humour, dans son livre comme dans la vie, Antoine nous propose – à la sauce « Maître Yoda » – des réponses et des solutions opérationnelles, fruits de son observation aiguisée des acteurs et ressorts de l'entreprise et de son expérience personnelle dans une start-up puis dans notre multinationale.

Mon seul regret, c'est qu'Antoine n'ait pu écrire ce livre il y a quatre ans.

Ma joie, c'est de savoir que tous les lecteurs vont avoir l'immense chance de pouvoir s'inspirer de ses riches observations et conseils pour les guider dans leur projet capital pour l'avenir de leur entreprise.

Jacques de Peretti
PDG AXA France

Copyright © 2018 Dunod.

Introduction

Triste destin que le vôtre (le nôtre) ! Vos faits, gestes et pensées forment la trame d'un récit continu, que vous ne contrôlez plus. Celui-ci est une succession de 0 et de 1, stockée, non sur des étagères, mais dans le *cloud*, aux quatre coins de notre belle planète. Bienvenue dans l'ère digitale ! Avant, vous décidiez lesquelles de vos data personnelles dévoiler, la plupart du temps en renseignant des formulaires de format papier. Le logo de l'entreprise apparaissait en haut à gauche du document, votre signature en bas de la page. Le contrat était clair.

> **Nous avons tendance à surestimer le changement à deux ans… et le sous-estimer à dix ans. Ne cessez pas d'accélérer !**
>
> **Bill Gates**

Aujourd'hui, vos pensées, ce sont des requêtes sur Google, vos désirs, des clics sur des liens. Sur écran, tout est traçable, tout est tracé. La transmission de vos data personnelles a explosé les frontières des formulaires. À présent, elle est continue et largement incontrôlée. Vous voici (vous et vos data) projetés sur la place publique. Qui, plus que d'un forum romain, a les allures d'un gigantesque supermarché.

Ce grand dévoilement des data personnelles, vos clients, vos collaborateurs le vivent avec la même intensité. Il fait le lit naturel des grands acteurs digitaux, californiens pour la plupart (Google, Facebook, Apple…) Leur promesse ? Elle a le mérite de la clarté : *Nous collectons vos data personnelles, de façon indolore, tout le temps, sans exiger de vous*

© Dunod - Toute reproduction non autorisée est un délit.
Copyright © 2018 Dunod

le moindre effort… pour les revendre à des annonceurs, sous la forme de publicités pertinentes et ciblées, adaptées à vos besoins. L'abandon de vos data est la condition de la gratuité des services que nous vous proposons, et que vous plébiscitez d'ailleurs au quotidien. À tout de suite ! Reconnaissez que le deal est attractif. Combien d'entre nous seraient prêts à payer Google, pour chaque requête effectuée sur son moteur de recherche ? Facebook, pour chaque photo de ses vacances publiée ? Entreprises privées, institutions publiques, *start-up*, associations… il n'est plus possible de reculer : apportez, avec vos équipes, votre réponse à ce grand dévoilement !

Un chiffre pour démarrer (vous l'avez sûrement déjà entendu, mais ça fait toujours bien de commencer par un gros chiffre) : en 2010, nous produisions déjà, en deux jours, autant de data que l'humanité dans toute son histoire. Le premier calculateur électronique Colossus, utilisé outre-Manche pendant la Seconde guerre mondiale, qui pouvait composer avec 5 000 caractères différents par seconde, fait déjà figure de dinosaure.

Autour de vous, une foultitude de concepts est relayée par les consultants, désormais évoqués jusque dans les journaux télévisés (quel est le pire ?). Ils sont tous plus importants les uns que les autres, mais peu compris. *Big data*, *datalake*, *Data Scientists*, algorithmes, *cookies*, *machine learning*… Vous en trouverez dans ces pages les définitions. Notre enjeu ? Donner du sens, un cap à ce mouvement de fond. La data, ce n'est pas seulement une nouvelle donne informatique. C'est un nouvel environnement économique, social et culturel. Quelle que soit votre taille, vous devez aider votre entreprise à y trouver une place. Puisse ce petit manuel, en vous accompagnant à chacune des étapes, vous aider sur ce chemin.

Copyright © 2018 Dunod.

1 CHAPITRE Définir votre stratégie

Prenez un peu de recul. Peu de ruptures technologiques vous obligent à repenser votre stratégie. Pensez au large mouvement d'informatisation des entreprises, dans l'après-guerre. Ou à celui de la digitalisation, depuis une vingtaine d'années. Pas si nombreuses, au bout du compte !

Eh bien c'est le cas, également, de l'irruption des data. Votre entreprise ne peut y rester indifférente. Il n'y a pas de neutralité possible. C'est votre stratégie globale qu'il faut éprouver. Demain, tout (nous assumons, tout !) produira de la data. Soyez-en persuadés. Et posez-vous la question cruciale : *Demain, à quoi mon activité doit-elle ressembler pour tirer parti de ce « tout data » ?* Étape indispensable avant de démarrer l'aventure : fixer le cap !

> **Mieux vaut prendre le changement par la main, avant qu'il ne vous prenne à la gorge !**
>
> **Winston Churchill**

Copyright © 2018 Dunod.

Copyright © 2018 Dunod.

Une étude du cabinet IDC, qui évalue la maturité du marché français sur le sujet, montre tout le chemin encore à parcourir : si 70 % des entreprises considèrent leurs data comme un actif stratégique, seuls 5 % des data internes qu'elles ont sous la main sont analysées.[1]

1 | LA DATA, POUR MIEUX FAIRE !

> ***EN VOILÀ UNE BONNE QUESTION !*** Comment exploiter mes data, pour rendre plus performante mon activité ?

En utilisant mieux et davantage vos data, vous rendrez votre entreprise plus efficace et améliorerez l'ensemble de ses procédures. Qu'est-ce qu'une procédure ? C'est une action, ou une succession d'actions, qui se répète, et qui obéit donc à une règle, plus ou moins explicite, plus ou moins formalisée. Cela touche toutes vos activités : achats, ressources humaines, distribution, production… Plus la procédure est automatique, plus la règle est formalisée, et plus le recours à la data est important. Prenez l'exemple d'une chaîne d'assemblage industrielle, ou de l'approvisionnement des rayons « liquides » d'une grande surface. Davantage de data (et de qualité !), c'est plus d'efficacité, de façon quasi-automatique.

Fixez donc une première ambition à votre entreprise : mieux faire, en ayant un recours astucieux à la data. N'hésitez pas des mois avant de vous lancer. La démonstration du retour sur investissement, lorsqu'il s'agit d'optimisation, est chose aisée. Suivez en cela vos ambitions de digitalisation ou de dématérialisation, souvent conduites en parallèle. Le point commun ? L'optimisation des procédures. Prenez un courrier papier reçu et lu à son domicile par votre client. Cela ne produit pas de data (aucune information formalisée, et partagée, si ce n'est la réception du courrier).

1. Cabinet FDC : *Observatoire de la donnée*.

© Dunod - Toute reproduction non autorisée est un délit.
Copyright © 2018 Dunod.

A contrario, en dématérialisant le courrier, l'e-mail, qui le remplace, produit (lui !) de la data : le clic d'ouverture, le temps passé, jusqu'au passage prolongé de la souris de votre client sur tel ou tel passage de l'e-mail… Rapprochez d'ailleurs et autant que possible les équipes chargées de la data et les équipes digitales. Les deux vont de pair. Chaque progrès dans la digitalisation est un pas vers l'automatisation… et donc un recours plus intensif à la data. Et là, faites fi des data externes ! Vos data internes permettent déjà de faire une vraie différence dans l'optimisation de vos procédures. En résumé : ne tergiversez pas ! Chacun de vos collaborateurs doit se poser la question *Comment mieux faire mon travail avec plus de data ?* Il s'agit, ici, d'excellence opérationnelle, pour ne pas dire d'hygiène opérationnelle (le terme n'est pas beau).

McKinsey Global Institute a analysé la capacité des secteurs industriels à valoriser leurs data, en fonction de quatre critères, qu'il est utile de rappeler[1] :

- **Le volume de data internes** dont vous disposez (vos gisements de data).
- **La variabilité de la performance** (calculée sur le résultat opérationnel). Un écart important entre les meilleurs et les moins bons performeurs de votre secteur d'activité est l'indication d'une forte opportunité à faire mieux *via* la data.
- **La volatilité de votre secteur**. À savoir, la fréquence à laquelle varient les entreprises les plus performantes et les entreprises les moins performantes de votre secteur.
- **La taille de votre marché** en termes de clients, de fournisseurs, et de transactions. Plus la taille est importante, plus vos data à utiliser sont nombreuses.

Pour fixer le bon niveau d'ambition, il vous faut connaître les data disponibles dans votre organisation et en posséder la cartographie,

1. McKinsey Global Institute, *Big data: The next frontier for innovation, competition, and productivity*, 2011.1

Copyright © 2018 Dunod.

par grande typologie : celles qui ont trait à vos clients, à la finance, à la production… Puis, il s'agit d'identifier vos gisements de data, exploités ou non. Ne vous lancez pas dans un long (et infini !) inventaire à la Prévert de vos data disponibles. Partez plutôt des grands bouleversements actuels, qui traversent les marchés et les entreprises. Des nouvelles data, et en quantité, sont produites dans leur sillage.

LA DISPARITION PROGRESSIVE DU PAPIER

Le remplacement du papier par des supports digitaux (e-mails, pages Web, formulaires…) permet une collecte plus complète et fiable des data, avec davantage de data structurées, donc plus facilement réutilisables. C'est d'ailleurs toute la beauté des formulaires digitaux. Finie l'écriture manuscrite, qui nécessite de l'interprétation (de l'humain !) et comporte des risques importants d'erreur. En même temps qu'il remplit votre formulaire digital, votre client structure l'information communiquée. Et vous pouvez contrôler la qualité de la data, en temps réel. Un oubli de code postal, une erreur de frappe sur le nom d'une rue ? Un contrôle de surface, à même le formulaire en ligne, peut détecter l'erreur et inciter le client à rentrer dans la « bonne case ». Chez vous, prêtez donc attention à tous les endroits où le papier est en train de disparaître. Et posez la question fatale : *Qu'est-ce que nous avons récupéré, en plus, comme data ?*

QUEL AFFREUX JARGON !
DATA STRUCTURÉES

Un ensemble de data est dit structuré lorsque toutes les valeurs présentes ont un lien logique entre elles, connu *a priori*, et qu'elles appartiennent à des catégories bien précises, dont le nombre et les définitions sont connus. Prenez l'image d'un tableau Excel, dont à la fois les lignes et les colonnes (au nombre limité) auraient des titres cohérents,

■■■

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

intelligibles et pérennes. Eh bien, il s'agit d'un tableau de data structurées. Un jardin à la française, pensé comme tel, avant d'y planter les arbres. *A contrario*, des data non structurées, ça correspondrait à un tableau Excel imaginaire, avec un nombre illimité de colonnes et de lignes, dont beaucoup d'entre elles n'auraient pas de titres, et dont le lien logique entre les valeurs ne serait pas connu.

LA CROISSANCE DE LA DISTRIBUTION DIRECTE

Les marchés des particuliers comme des entreprises sont touchés par l'essor du e-commerce. C'est une formidable opportunité pour les producteurs. Ils dépendent de moins en moins de leurs distributeurs, et peuvent (enfin !) développer une connaissance de leurs clients, et de leur consommation. Pour vous, la question (pas aussi simple qu'elle en a l'air !) devient : *Comment adapter mon offre par rapport à la façon dont elle est consommée par mes clients ?*

ILS L'ONT FAIT !

L'ajustement d'une offre à la façon dont elle est consommée ? Un phénomène qui touche des secteurs variés.

Dans l'automobile, l'usage des configurateurs sur les sites Web des marques (les choix d'options notamment) oriente des décisions de production industrielle.

Dans les médias, le *zapping* des internautes sur tel ou tel passage d'une vidéo en *replay* oriente les choix éditoriaux de la rédaction.

Dans le voyage, le nombre de visites sur les sites de réservations de billets d'avion est intégré comme une variable du prix affiché.

Dans le monde de l'édition, *via* l'adoption des liseuses numériques, les éditeurs seront demain tentés d'infléchir le contenu même des ouvrages de leurs écrivains.

Copyright © 2018 Dunod.

Enfin, citons l'industrie cinématographique. Certaines *start-up* affirment pouvoir prévoir le succès d'un film, sur la base de son scénario… et donc d'éviter l'accident industriel ! L'algorithme s'appuierait sur de nombreux critères, comme le thème, le niveau de violence, les lieux et la fréquence des effets spéciaux. La *start-up* israélienne Vault, par exemple, prétend pouvoir prédire le succès d'un film en salles avec un taux de précision de 70 % !

IL TÉMOIGNE !

BARNABÉ ANGLADE, UBISOFT

Voilà plusieurs années que l'industrie du jeu mobile a opéré un tournant vers les data. Celui-ci est allé de pair avec l'émergence du modèle dit « *Free-To-Play* » qui désormais domine largement en termes de revenus, à comparer au modèle traditionnel dit « *Premium* ».

Là où il s'agissait initialement de faire « le meilleur jeu possible » afin que la qualité dudit jeu justifie qu'on veuille l'acheter, il s'agit à présent de proposer une expérience prompte à garder un maximum de joueurs sur la durée, dans l'idée qu'une partie d'entre eux puisse à terme effectuer une ou plusieurs transactions au sein d'un jeu que n'importe qui peut télécharger gratuitement.

En d'autres termes, l'ancien modèle dominant revenait à opérer la meilleure adéquation entre le prix d'entrée du jeu, sa qualité et son public. À présent, il s'agit de rentabiliser une production qui peut virtuellement rapporter zéro, un utilisateur pouvant en profiter sans débourser le moindre centime. Pour ce faire, il revient aux designers et aux développeurs d'accoucher de mécaniques promptes à assurer un maximum de conversions parmi cette population. Il faut donc pouvoir déterminer les comportements des joueurs. Sont-ils assez nombreux au terme du tutorial ? Cette offre est-elle assez alléchante ? Combien de temps passent-ils en moyenne sur l'application ? Le niveau 3 est-il trop facile ? Difficile ?

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

■■■

Ce nouveau modèle, souvent plus proche du Web marketing et du e-commerce que du jeu traditionnel, se heurte encore à des questions éthiques, certains ayant parfois l'impression que la qualité de l'œuvre collective passe désormais derrière la nécessité d'optimiser des objectifs chiffrés. Ces questions demeurent plus que jamais légitimes. Et parce qu'il n'y a pas de modèle financier unique, il convient de revenir en permanence au joueur. Quelle est la qualité d'expérience d'un joueur qui décide de jouer sans payer ? Quelle est la valeur de l'offre proposée ? Du temps ? Du confort ? Des services additionnels ?

Dans le cadre d'un marché de masse, un transfert s'opère entre l'œuvre qu'on diffusait comme « finie » et l'idée qu'à présent, un jeu est avant tout la mise en place d'un service. Sauf que l'idée de service ne doit pas effacer celle du jeu en tant que tel.

Fort heureusement, la maturité grandissante du public face à ce genre de modèle permet de remettre au centre la raison pour laquelle beaucoup s'orientent vers cette industrie : il ne s'agit pas de forcer des joueurs à la conversion, mais de leur proposer une expérience qui mérite qu'ils y investissent leur temps et, si une des offres proposées leur paraît valable, un peu de leur argent, au risque de perdre ces joueurs si la transaction n'est pas jugée satisfaisante.

Si la data est désormais une nécessité pour qui veut mesurer la qualité d'un jeu face à des joueurs, elle ne dispense pas de se poser la seule question qui doit selon moi guider le *game design* : quelle est la valeur de l'expérience proposée ?

L'EXPLOSION DES DATA EXTERNES

Ne vous attendez pas à quelque chose de renversant, mais prenez régulièrement le pouls des fichiers externes disponibles sur le marché. Le mouvement de fond est bien là : vos systèmes informatiques, autrefois fermés, s'ouvrent progressivement à des partenaires externes, *via* les fameuses API (pour le détail, rendez-vous à la partie « Le Grand Soir technologique,

Copyright © 2018 Dunod.

c'est ce soir!» p. 67). Ne soyez pas naïfs : dans la grande majorité des cas, l'accessibilité des data est payante. *Open Data* rime (trop souvent) avec *Paid Data*. Les data externes peuvent être très proches de vous. Prenez vos prestataires ou vos partenaires, notamment marketing. L'échange ou la vente de data ne font pas partie de vos engagements contractuels ? C'est dommage, car c'est une belle opportunité pour enrichir le partenariat. Demandez-vous régulièrement : *Quelles data intéressantes sont disponibles hors les murs ?*

MAIS COMMENT ÇA FONCTIONNE, CONCRÈTEMENT ?

Les gouvernements ont également pris le tournant de l'*Open Data*, plus ou moins rapidement. Le site www.data.gouv.fr liste l'ensemble des data publiques disponibles. Exemple iconique : l'INSEE, qui rassemble de précieuses data, issues du recensement de la population, et qu'elle rend disponibles gratuitement.

Sur la base des data publiques, la *start-up* VroomVroom a réussi à présenter les auto-écoles sur la base de leur taux de réussite, exercice de transparence jusqu'ici jamais réalisé.

La SNCF a rendu accessible les trajets de ses trains, *via* une application publique : https://data.sncf.com.

Des data simples, anodines en apparence, ont un fort impact sur votre activité. Vous en doutez ? Prenez par exemple la météo. Le cabinet Climpact montre que 70 % de l'économie y est sensible : une baisse de 2 degrés de température au printemps, c'est une hausse de 30 % des ventes de fromages à raclette et une baisse de 8 % des ventes de sirop[1]. Tiens, qu'est-ce que ça changerait à vos activités, de prendre en compte la météo à venir ?

© Dunod - Toute reproduction non autorisée est un délit.
Copyright © 2018 Dunod.

1. www.climpact.com

LA DIGITALISATION DE VOS CLIENTS

C'est un fait (donc pas de débat possible), vous passez plusieurs heures par jour devant des écrans. Et malheureusement, le temps n'est pas extensible. Les réseaux bancaires voient moins de clients passer dans leurs agences. Les conseillers sont donc moins au courant de la vie de ces derniers, notamment de leurs fameux moments de vie. La bonne nouvelle ? Vous pouvez continuer à suivre vos clients lorsqu'ils sont en ligne. Leurs navigations digitales sont mesurées, et peuvent enrichir votre connaissance-client déjà existante, à condition de collecter les data digitales, et de les lier à vos autres data classiques. Des outils sont précisément là pour faire le lien entre l'identité digitale de vos internautes (le *cookie*), et leur identité classique (nom, adresse postale ou e-mail). Vous trouverez l'explication dans la partie « Les data digitales, tu chériras ! » (p. 116).

Vos collaborateurs sont témoins des grands mouvements évoqués ci-dessus. Votre activité, vos procédures, ne resteront pas identiques.

> ***EN VOILÀ UNE BONNE QUESTION !*** Qu'est-ce que ça change pour mes équipes ?

Ne laissez aucune équipe passer à côté de ce questionnement. Même celles qui en semblent (à première vue !) les plus éloignées.

Prenez la direction des ressources humaines. C'est le territoire par excellence de l'intuition, de l'humain, du particulier.

Google s'est lancé ce défi : réconcilier ressources humaines et data. Pour ce faire, l'entreprise a lancé en 2009 le programme Oxygène, avec deux ambitions. D'abord, prouver l'utilité du management à l'ensemble de l'organisation (et notamment aux ingénieurs, peu convaincus de l'utilité des organigrammes et des fonctions support). Puis, aider les managers à progresser. Signal significatif : la direction RH s'est rebaptisée *People Operations*. *Via* un travail data colossal, les équipes du programme ont prouvé la corrélation positive entre la performance d'une équipe et la qualité de son manager (bonne nouvelle, vous voilà rassuré !).

Copyright © 2018 Dunod.

Seconde étape : la systématisation de sondages fréquents, obligatoires auprès des équipes (les UFS, pour *Upward Feed-back Survey*) ont permis d'identifier les huit critères d'un manager performant, de les diffuser et d'accélérer l'accompagnement et la formation des managers les plus en difficulté. En moins de deux ans, le taux de réponse favorable médian collecté par ces fameux UFS est passé de 83 % à 88 %.

Ce qui est vrai pour le *management*, l'est aussi pour le recrutement. Une meilleure utilisation des data permet d'affiner la qualité de la sélection. Dans ses centres d'appel, Xerox a établi que les meilleurs candidats sont ceux qui sont actifs sur les réseaux sociaux, habitent près de leur lieu de travail et bénéficient de transports en commun de qualité. Intéressant, non ?

Avec quelques data, vous pouvez faire rapidement la différence. Citons ici le chercheur Daniel Kahneman. Dans son illustre ouvrage *Thinking, Fast and Slow*[1] (à lire immédiatement après avoir fini de dévorer celui-ci), il démontre les limites de l'intuition lors des entretiens de recrutement. Connaissez-vous l'effet de halo, auquel vous êtes sujet ? *Parce que ce candidat est souriant et sympathique, j'en déduis intuitivement qu'il est compétent.*

Et l'auteur de recommander une démarche plus objective, sur la base d'un questionnaire. Lisons-le (et taisons-nous) : *Supposons que vous devez engager un représentant commercial pour votre entreprise.*

Tout d'abord, sélectionnez quelques critères qui sont des conditions préalables au succès du candidat à ce poste (compétences techniques, personnalité engageante, fiabilité…). N'exagérez pas : six dimensions sont un bon nombre. Les critères que vous choisissez devront être aussi indépendants que possible. Et vous devez sentir que vous pouvez les évaluer de manière fiable, en posant quelques questions factuelles. Ensuite, faites une liste de questions pour chaque critère.

Ces préparatifs devraient vous prendre une demi-heure environ, un petit investissement qui peut faire une différence significative dans la qualité des personnes que vous recrutez.

1. Traduit en français : *Système 1 / Système 2 : les deux vitesses de la pensée*, Flammarion, coll. « Essais », 2012.

© Dunod - Toute reproduction non autorisée est un délit.
Copyright © 2018 Dunod.

Pour éviter l'effet de halo, dans l'entretien, vous devez collecter les informations sur un critère à la fois, en marquant chacun, avant de passer au prochain. Ne pas sauter. Pour évaluer chaque candidat, additionnez les six scores (évaluation de 1 à 5, de « très faible » à « très fort », par critère)…

Engagez-vous fermement à embaucher le candidat dont le score final est le plus élevé, même s'il y en a un autre que vous lui préférez. Résistez à votre volonté d'inventer de nouveaux critères pour changer le classement.

Toutes les grandes études académiques convergent sur cette promesse : vous trouverez un meilleur candidat si vous utilisez cette procédure, plutôt que de faire ce que les gens font normalement dans de telles situations, c'est-à-dire se présenter à l'entretien sans préparation… et le conclure par des paroles aussi hâtives que : « je l'ai regardé dans les yeux et j'ai aimé ce que j'ai vu ».

Allez jusqu'à appliquer l'un des critères de McKinsey mentionnés ci-dessus. Dans une même équipe, et pour une activité similaire, plus les performances annuelles varient d'un collaborateur à un autre, plus la data a du potentiel.

DES EXEMPLES

Picorez quelques idées (c'est gratuit !) :

Achat | Ayez recours aux data externes pour piloter la santé financière de vos fournisseurs et prestataires, et sécuriser l'approvisionnement.

Risque | Équipez de sondes et de capteurs votre chaîne de production, pour remonter en temps réel des risques de sécurité, et vous permettre d'agir en prévention.

Fraude | Corrélez un maximum de data possibles, pour identifier des comportements de fraude ou d'abus, chez vos clients.

Marketing | Captez, *via* les navigations digitales de vos clients, des signaux précoces d'attrition, pour vous permettre d'anticiper des résiliations (usage de votre espace-client, requête sur Google…)

Service-client | Identifiez les paires les plus performantes entre vos conseillers et vos clients, sur la base de leurs profils sociodémographiques, comme psychologiques.

Copyright © 2018 Dunod.

2 | LA DATA, POUR FAIRE AUTREMENT !

EN VOILÀ UNE BONNE QUESTION ! Comment développer de nouveaux revenus avec la data ?

Utiliser pour mieux faire, c'est déjà un challenge d'envergure et ambitieux pour beaucoup d'entreprises. Vous pourriez vous arrêter là. Mais poussez un peu plus la réflexion. Et si la data vous permettait de repenser complètement votre proposition de valeur. Quid de la monétisation, directe ou indirecte, des data en votre possession ? Et d'une nouvelle ligne de revenus ?

Dans ce cas, votre compte d'exploitation, demain, n'aura pas la même structure qu'aujourd'hui. Il est indispensable de fixer cette ambition en amont (si elle a un sens pour votre activité bien sûr !), car elle bousculera vos collaborateurs, et nécessitera de déployer une stratégie d'investissement autrement plus importante que celle évoquée dans la partie « La data, pour mieux faire ! » (p. 3). Vous projetez alors l'ensemble de votre entreprise sur de nouveaux territoires, à la conquête de nouvelles data !

Rassurez-vous (ou plutôt inquiétez-vous !), votre activité n'est pas forcément concernée. Un indice qui a son importance pour le savoir : avez-vous déjà en votre possession un gisement de data, déjà exploité ou exploitable ?

QUEL AFFREUX JARGON !

GISEMENT DE DATA

Cette expression désigne un volume important de data, qui répond à deux critères. D'abord, le fait que ces data vous soient exclusives (vous êtes le seul acteur à les posséder, ou presque). Puis, que vous en fassiez la collecte dans l'exercice naturel de votre métier (ou dans son prolongement).

■■■

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

Les gisements de data sont déjà plus nombreux que les gisements de pétrole : la localisation des communications pour les acteurs de la téléphonie, les profils de risque pour les assureurs, les recherches en ligne pour Google, les informations fiscales pour le Trésor Public… Bonne nouvelle ! La plupart des gisements de data restent inexploités. Et (plus grave peut-être), sous-estimés.

L'étude IDC, déjà citée, dit que 49 % des entreprises, par exemple, ne voient pas d'intérêt dans les data issues de la mobilité et des consommateurs…

ILS L'ONT FAIT !

Pour les acteurs du *trading* financier, la data a déjà tout changé ! L'expertise du *trader* est progressivement remplacée par des algorithmes sophistiqués, s'appuyant sur une infrastructure technologique de pointe, qui permettent de réaliser des transactions à très haute fréquence.

En 2015, Goldman Sachs sort du bois, et s'affiche au marché comme une *Tech Company*. Sa promesse ? Devenir le Google du risque financier (pas moins !). À savoir, collecter le maximum de data (transactions financières bien sûr, mais aussi événements politiques, sociaux, climatiques…), en un lieu unique (chez eux !), pour pouvoir tenir leur promesse client : *Certains clients portent des risques qu'ils ne souhaitent pas ; d'autres souhaitent des risques qu'ils ne portent pas : Goldman Sachs fait en sorte que les deux populations soient satisfaites*.

Il s'agit, à chaque demande d'un client, d'associer de façon automatique une recommandation qui satisfasse son appétit au risque. Sur ce chemin, le banquier new-yorkais a réuni les ingrédients nécessaires.

D'abord un recrutement massif d'ingénieurs, effort commencé dès les années 1990. Goldman Sachs compte aujourd'hui un tiers d'ingénieurs dans ses rangs, davantage, en nombre absolu, que Facebook.

Copyright © 2018 Dunod.

Puis, une exigence continue d'automatisation du conseil, en réduisant un maximum les interactions humaines. La suite de logiciels qu'ils ont développée a enregistré, en 2015, plus de 175 milliards de transactions, le double de l'année passée.

Enfin, une obsession de l'ouverture aux partenaires. Leur plateforme est ouverte, *via* des API (plus de détails dans la partie « Le Grand Soir technologique, c'est ce soir ! », p. 67) à l'écosystème de leurs clients, de leurs prestataires… mais aussi du régulateur !

Sa stratégie d'investissement en Venture Capital témoigne d'un appétit très aiguisé pour la technologie. Goldman Sachs a investi dans Uber, Pinterest, Dropbox… Comme une façon, au-delà de l'investissement financier, d'acculturer les collaborateurs Goldman Sachs à ces nouveaux *business*. Et, en retour, d'inciter ces derniers à développer sur sa plateforme technologique.[1]

Identifier un gisement de data est une chose. Se donner l'ambition d'en tirer des revenus en est une autre. Préparez-vous à tâtonner, à expérimenter pour déterminer le bon modèle économique. Mais votre intuition initiale (et celle de vos équipes) sur la valeur de votre gisement de data est indispensable. Décider de faire évoluer votre modèle économique tient du *Big Bet* (comme disent les consultants). Il faut y réfléchir à deux fois, avant de se lancer ! Citons deux façons différentes de monétiser vos data, pour en faire une nouvelle ligne de revenus.

VENDEZ DIRECTEMENT VOS DATA

Pour ce faire (et être rentable), vous devez réunir deux conditions. D'abord, collecter un volume important de data. Puis, que la valeur de ces data (c'est un paradoxe !) soit assez moyenne. Si chacune de

1. www.youtube.com/watch?v=VF6DrX9H0Ug&feature = youtu.be

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

vos data, prise à l'unité, avait une très forte valeur, vous auriez tort de la vendre directement. Ce serait plus rentable de la transformer, pour l'intégrer dans une proposition de valeur plus globale (à plus forte marge). Dans cas, rendez-vous au paragraphe suivant, « Enrichissez votre proposition de valeur ! » (p. 18). Quel intérêt, pour un assureur par exemple, de vendre directement les profils de risque de ses clients à d'autres acteurs, alors qu'il peut s'en servir pour enrichir (et personnaliser) ses propres produits d'assurance ? Trop faible !

Les cas les plus évidents de monétisation directe, vous les trouverez plutôt du côté de la publicité. Prenez Google ou Facebook. Ce sont des services gratuits pour nous autres utilisateurs. La vente de nos data, sous la forme de profils publicitaires, est la rançon à payer pour cette gratuité. Google commercialise vos requêtes sur son moteur de recherche ; Facebook votre profil sociodémographique et vos goûts… Même des grands médias plus historiques (sans aucune connotation péjorative) s'y mettent. Le site lemonde.fr ne monétise plus seulement des espaces présents à côté de tel ou tel article, mais également les profils de ses lecteurs, en fonction de leurs centres d'intérêt.

LeBonCoin est désormais en capacité de vendre aux annonceurs les profils des amateurs de tel ou tel véhicule, en fonction de leurs navigations sur ses petites annonces…

Ce métier de monétisation de la data, c'est aussi celui d'acteurs plus institutionnels, comme La Poste. C'est plus de 40 millions d'adresses postales, 26 millions d'e-mails et 19 millions de numéros de téléphone que vend (à qui veut bien les acheter !) sa filiale data, MédiaPost.

Ne restez pas cantonnés à la publicité ou au marketing direct. Avec l'émergence et le développement de la voiture autonome, de nombreux constructeurs automobiles envisagent de vendre les data techniques et de conduite des véhicules : aux assureurs, aux acteurs de la mobilité…

L'État lui-même, *via* l'ouverture des data publiques, aura la tentation, demain, d'en commercialiser certaines. Rêvons : un jour, l'État ne

Copyright © 2018 Dunod.

lèvera plus l'impôt, mais tirera ses revenus de la monétisation des data collectées sur les citoyens… Attention : ne vous précipitez pas sur ce chemin ! La monétisation directe des data personnelles est une activité hautement régulée, et encore peu comprise par l'opinion publique française. Contrairement aux populations anglo-saxonnes (et surtout américaines), plus réceptives.

ILS L'ONT FAIT !

La grande distribution recèle de formidables gisements de data. Les paiements des clients en font partie. La tentation est grande, pour les acteurs du secteur, de vendre directement, à côté de leur *business* classique, les profils de leurs consommateurs, à des fins publicitaires. C'est le cas de Carrefour, client depuis 2011 d'Axciom (éditeur technologique américain). Le principe : reconnaître sur le Web l'ensemble des clients Carrefour, *via* un processus de « cookification » (expliqué dans la partie « Les data digitales, tu chériras ! », p. 116), pour proposer aux annonceurs de les cibler avec des messages publicitaires.

Par exemple, pour un constructeur automobile, ça peut être particulièrement intéressant de pouvoir cibler, sur Internet, les familles avec deux enfants s'apprêtant, au vu de leur consommation, à en avoir un troisième.[1] Carrefour, mais aussi (et surtout !) Amazon, sont devenus des régies publicitaires à part entière. En 2017, la firme américaine dégageait (déjà !) plus d'un milliard de dollars de revenus publicitaires. Son gisement de data ? Pareil, sa base de connaissance des consommateurs, identifiés comme tels, qui représente de l'or pour les annonceurs soucieux d'optimiser leurs investissements marketing.[2] Enfin, disons un mot d'une initiative récente, réunissant à la fois des groupes média et des distributeurs français (Les Échos, Lagardère, SoLocal, SFR, Fnac/Darty…).

■■■

1. « Carrefour Média connecte les données pour les marques », Alliancy, 27/04/2016.
2. « Amazon s'attaquera-t-il aux agences de pub ? », *L'ADN*, 7/03/2017.

© Dunod - Toute reproduction non autorisée est un délit.
Copyright © 2018 Dunod.

Ils projettent de mutualiser leurs data pour enrichir les profils de leurs internautes respectifs. Et de concurrencer Google et Facebook. Le produit s'appelle Gravity. Souhaitons-leur bonne chance ! La copropriété, en termes de data, mêlée à des enjeux publicitaires, et avec des copropriétaires aux maturités technologiques pour le moins contrastées… ce ne sera pas une mince affaire ![1]

ENRICHISSEZ VOTRE PROPOSITION DE VALEUR

Si votre gisement de data est de très haute valeur, ne le vendez pas de façon sèche. Intégrez-le dans une proposition de valeur plus globale. Proposez, par exemple, des nouveaux services à vos clients, sur la base des data collectées.

Une question simple, pour démarrer votre réflexion : *Comment mieux restituer les data collectées sur mes clients en retour ?* Naturellement, plus votre entreprise collecte de data, plus vous êtes concernés. Pensez aux banquiers et aux assureurs. Commencez par ces derniers. Pendant des décennies, ils ont collecté des data sur les profils de risque de leurs clients, pour ajuster la prime d'assurance payée. Leur promesse ? *Nous serons là, en cas de difficultés*. Au-delà de cette promesse (fondamentale), quels autres services les assureurs rendront-ils à leurs clients, en échange de leurs data personnelles ? Quelques illustrations de conseils qu'ils pourraient prodiguer à leurs clients (plus ou moins « science-fictionesques » !).

En épargne : *Les clients avec votre profil partent à la retraite, en moyenne, à tel âge, et avec tel montant de retraite. Préparez-vous.*

En habitation : *Vous vous apprêtez à emménager dans un immeuble sujet, depuis deux ans, à de nombreux dégâts des eaux, négociez à la baisse le prix d'achat.*

1. « Pub en ligne : les médias français font front commun. », *Les Échos*, 04/07/2017.

Copyright © 2018 Dunod.

En santé : *Au vu de vos consommations récentes de médicaments, vous devriez faire un dépistage de telle pathologie.*

L'exemple de la santé n'est pas anodin. En France, la loi Evin interdit en effet d'utiliser les data médicales pour personnaliser le coût de la mutuelle. Elles serviront davantage à enrichir l'offre existante de nouveaux services, pour améliorer le parcours de soin des patients.

Prenez maintenant le cas des banquiers : pendant des années, qu'ont-ils fait des data de transactions de leurs clients ? Presque rien, si ce n'est identifier (et facturer) les découverts et repérer les forts encours, pour proposer de l'épargne.

Demain, ils pourraient (en vrac) aider à mieux consommer : *Ne vous arrêtez pas à cette station-service, une autre, à quelques kilomètres, est moins chère.*

À mieux prévoir : *Attention, parti comme vous êtes ce mois-ci, vous aurez un vilain trou de trésorerie à la fin du mois.*

Ou encore aider à mieux épargner : *Si tu mettais chaque mois tant d'euros sur ton livret, tu gagnerais tant d'euros pour ta retraite, dans cinq ans*.

ILS L'ONT FAIT !

Quittons le secteur financier pour l'industrie. Rolls Royce a également opéré, ces dernières années, une vraie révolution de modèle économique. D'un statut de fabricant de moteurs d'avion, l'entreprise est devenue fournisseur de services.

Elle vendait un moteur en *one-shot* à ses clients, puis faisait sa marge sur les réparations et la maintenance. Désormais, la firme britannique commercialise ses produits à l'usage, c'est-à-dire aux heures de vol des avions. Et dans l'abonnement souscrit par le client sont inclus les coûts de maintenance (les réparations comme les remplacements).

■■■

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

Pour rendre ça possible, Rolls Royce a dû mettre en place un pilotage constant et robuste du fonctionnement de ses moteurs, une fois livrés aux clients. Et donc procéder, en interne, à une révolution dans la collecte, le traitement et l'utilisation de ses data. Les récentes ruptures technologiques, notamment la connectivité des objets a aidé.

L'entreprise s'est posée la question existentielle : *Quel est mon métier ?* Important ! La réponse apportée n'a pas été *Fabriquer des moteurs d'avion*, mais bien *Permettre aux avions de voler*.

D'où l'ambition data déclinée naturellement, dans un second temps.

Enfin, évoquons un tout autre secteur, celui de l'agriculture. La politique d'acquisition externe du leader mondial des semences, Monsanto, dit bien sa stratégie. Il s'agit d'une succession de rachats d'entreprises spécialisées dans la data et la prédiction (*The Climate Corporation*, *Precision Planting*, *Solum*…). L'enjeu de Monsanto ? Se réinventer comme une entreprise de services, et non plus comme un seul fournisseur de semences. Leur métier ? Accompagner les agriculteurs dans l'optimisation du rendement de leurs cultures, en leur apportant un conseil approprié (*via* la data !)

ILS NE L'ONT PAS FAIT…

Un acteur majeur de l'économie et de la société ? Qui en connaît le plus sur nous (après Google !) ? Vous voyez ? Nous parlons… de l'État !

Beaucoup de data collectées, pour quels services en échange ?

Si l'État s'est engagé sur la voie de l'*Open Data*, le mouvement est bien timide. Il n'est pas en train de se renouveler, il optimise, sans poser la question de son métier. Pourtant, *via* ses data, il pourrait se penser (puis se transformer) en « État-plateforme », en s'appuyant davantage sur des partenaires, privés comme bénévoles, pour assurer et enrichir nombre de missions de service public.

Copyright © 2018 Dunod.

Pour ce faire, il mettrait à disposition de tous une plateforme technologique ouverte, rassemblant des data, internes comme externes, et pouvant exécuter des transactions.

Des exemples plus concrets ? Des *start-up* pourraient se brancher aux data de Pôle emploi et proposer aux chômeurs des offres d'emplois plus ajustées, issues d'algorithmes plus puissants.

Les innombrables espaces client Web pourraient interroger l'État pour certifier l'identité de leurs clients sans les obliger à renseigner des formulaires.

Les agendas de l'ensemble des institutions publiques (dont les hôpitaux) seraient rendus accessibles pour des prises de rendez-vous en ligne.

Une application mobile d'entraide entre voisins pourrait être animée par les municipalités.

Les citoyens pourraient signifier aux autorités les nids de poule prioritaires à combler sur les routes mal entretenues.

Enfin (lâchons-nous !), Amazon pourrait être chargé de collecter la TVA sur la vente en ligne des commerçants présents sur sa plateforme…

Le chemin est encore long, mais qu'attend l'État pour fixer un horizon positif, qui donnerait sens à tous les efforts de productivité attendus de ses fonctionnaires (et pas prêts de s'arrêter) ? L'État français pourrait (une fois n'est pas coutume) s'inspirer de l'un des plus petits des États de l'Union européenne, l'Estonie.

L'ensemble des services aux citoyens y est entièrement digitalisé. Et l'Estonie d'œuvrer, à l'échelle européenne, pour qu'aux côtés des quatre libertés de circulation fondamentales (celle des biens, services, capitaux et individus), en figure une cinquième, celle de la data.

En conclusion : commencez par vous poser la question existentielle *Quel est mon métier ?* Et apportez-y une réponse large, centrée sur le client, et indépendante de la façon opérationnelle que vous avez (aujourd'hui) d'apporter de la valeur. Si vous fabriquez des baignoires,

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

ne répondez pas *Fabriquer des baignoires*, mais bien *Aider les clients à rester propres*. Si vous êtes le Trésor Public, ne répondez pas *Lever l'impôt auprès des contribuables*, mais bien *Permettre le bon fonctionnement des services publics*… À condition d'avoir une vision élargie de votre métier, votre entreprise pourra trouver l'énergie de se transformer radicalement. Vous vivez dans un environnement économique en constante recomposition. C'est une condition de survie. Et un recours plus intensif, et différent, à vos data, vous aidera à poursuivre votre métier, autrement.

> **Les opportunités se multiplient lorsqu'elles sont saisies.**
>
> **Sun Tzu**

Copyright © 2018 Dunod.

3 | LA DATA, POUR NE PAS DISPARAÎTRE !

EN VOILÀ UNE BONNE QUESTION ! Comment maintenir, et développer, le lien de confiance avec mes clients ?

À vrai dire, vous auriez pu zapper les deux parties précédentes. Déclarer que la data ne vous servira ni à faire mieux, ni à faire autrement. Et être tenté de ne rien faire, tout simplement. C'est votre choix.
Notre devoir est de vous avertir : en termes de data, l'immobilisme n'est pas une option. Vous seriez, malheureusement, rattrapés par la patrouille. En l'occurrence deux patrouilles, aux *modus operandi* et aux habitudes vestimentaires bien différents : les régulateurs et les *hackers*.

> **Movimiento es la vida !**
>
> **Brad Pitt (World War Z)**

LES RÉGULATEURS

En mai 2018, l'Union européenne bascule dans une nouvelle ère, avec la mise en application du règlement européen GDPR (*General Data Protection Regulation*). Il s'applique à toutes les entreprises qui collectent, traitent et stockent des data personnelles, qui permettraient d'identifier directement ou indirectement un individu. Pour vous, pas de politique de l'autruche possible ! Le règlement bouleverse votre entreprise, tant sur sa gouvernance, sa gestion des systèmes informatiques que ses procédures. L'article vingt porte, à lui seul, un principe quasi-révolutionnaire. Nous le détaillons dans la partie « De la réglementation tu te préoccuperas ! » (p. 103). Citons (en guise d'apéritif !) quelques-unes de ses mesures emblématiques.

- Garantir, à la demande de votre client, que toutes ses data seront supprimées de vos systèmes informatiques, sous un délai de trente jours.
- Mettre à sa disposition, s'il le demande, toutes les data personnelles qui le concernent sous un format lisible et interprétable par une machine… et aller jusqu'à les transférer à une autre entreprise

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

(comme dans le cas d'une portabilité de numéro chez les opérateurs téléphoniques).

- En cas de cyberattaque, avertir les clients concernés sous soixante-douze heures…

Conclusion : gare à l'insouciance ! Jusque dans la conception même de vos offres et produits, pensez à la protection des data personnelles *by design*, à savoir dès le départ. Le régulateur vous y encourage aujourd'hui, et demain (en mai 2018) vous y obligera.

LES HACKERS

Posons d'abord le décor. En 2016, 80 % des entreprises françaises ont constaté une cyberattaque lors des douze derniers mois[1]. 500 millions, c'est le nombre de comptes personnels Yahoo ! piratés en 2016. 800 000 euros, c'est le coût moyen pour se remettre d'une cyberattaque[2]. Enfin, 200 000, c'est le nombre de machines touchées, en une journée, par le (tristement fameux !) virus *WannaCry*, le vendredi 12 mai 2017…

Qu'est-ce à dire ? Que les cyberattaques révéleront, tôt ou tard, le peu d'intérêt que vous portez aux data, en exposant au grand jour vos écarts par rapport aux principes de régulation.

L'actualité en témoigne : aujourd'hui, les CEO « tombent » davantage pour des raisons de fraudes, de failles de sécurité que pour des mauvais résultats…

Pour finir de se convaincre que la cybersécurité est l'affaire de tous, citons l'exemple de la Réserve Fédérale américaine. Entre 2011 et 2015, elle a enregistré plus d'une cinquantaine d'intrusions malveillantes. En 2016, c'est près d'un milliard de dollars qui ont failli disparaître, suite

1. Baromètre Cesin : 80 % des entreprises françaises ont constaté au moins une cyberattaque dans l'année, 26/01/2017.
2. « Cybersécurité : 5 chiffres-clés à connaître », Microsoft Ideas, 30/08/2016.

Copyright © 2018 Dunod.

à une cyberattaque particulièrement intrusive, finalement contenue à une perte de 80 millions de dollars. Vous l'aurez compris, personne n'est à l'abri !

La version négative : si vous n'êtes pas vigilant sur la gestion de la data, vous vous exposez à des difficultés existentielles.

La version positive : profitez de la saine pression des régulateurs comme des *hackers* pour bâtir, maintenir, et développer une relation de confiance avec vos clients, en montrant patte blanche, et en faisant du respect des data personnelles un engagement fort de votre entreprise.

Mais attention : prendre la parole pour afficher votre vertu vous oblige. Vos clients doivent sentir un alignement entre ce positionnement data et la proposition de valeur de votre entreprise, les valeurs véhiculées par votre marque. Gare au *privacy washing* ! Et ne sous-estimez pas votre client : il sentira vite la supercherie. Peu d'entreprises visent juste en la matière. Citons quelques rares exceptions : La Poste, MAIF, ou encore Microsoft, d'une autre façon.

En termes de *Data Privacy*, le mimétisme a joué à plein. Récemment, la plupart des grandes entreprises ont pris la parole. Résultat : une foultitude de manifestes et autres chartes sur les data personnelles, sans vrai relief ni engagement saillant. Goûtez-en, si l'envie vous prend, un florilège ![1] Ces engagements ne font pas de mal en soi. Mais ils ne suffisent plus pour se différencier. Soyez plus ambitieux pour votre entreprise !

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

1. Orange, « Charte protection des données personnelles et de la vie privée », https://bienvivreledigital.orange.fr/mes-donnees-mon-identite/charte-protection-des-donnees-personnelles-et-de-la-vie-privee
Crédit agricole, « Charte des données personnelles », www.credit-agricole.fr/nos-engagements/charte-des-donnees-personnelles.html

Posez-vous plutôt la question de la *value for data* perçue par vos clients. Quel écart entre le volume de data que nous collectons sur nos clients et l'utilité perçue par ces derniers, en retour ? Si l'écart est trop important, c'est votre proposition de valeur qui est en cause. Retournez vite à la partie « La data, pour faire autrement ! » (p. 13).

Copyright © 2018 Dunod.

QUEL AFFREUX JARGON !

VALUE FOR DATA

Le concept est assez simple (mais il fallait l'inventer). C'est l'équivalent du principe de *value for money* (*En tant que client, qu'est-ce que j'ai en retour du prix que je paie* ?), mais appliqué cette fois-ci à la data (*En tant que client, qu'est-ce que j'ai en retour des data que je vous transmets ?*). Et, en la matière, la donne a changé. Pendant des années, vous avez collecté, *via* des formulaires (tous aussi barbants les uns que les autres), de nombreuses data sur vos clients, et ce sans avoir à rendre compte directement de l'usage que vous en faisiez en retour.

Désormais, le grand public a conscience de la valeur de ses data personnelles. Il veut en tirer des avantages, et davantage, à l'image des programmes de fidélité : des data de consommation contre des réductions. Plus récemment, prenez l'exemple de Google. Le moteur de recherche californien a construit son succès sur le principe du *free for data* : nos data personnelles, à vocation publicitaire, en contrepartie de l'accès gratuit à ses services. Reconnaissez que le deal est attractif. Vous seriez prêt, vous, à payer Google pour chaque requête effectuée sur son moteur ?

Vous avez recours à des algorithmes ? Affichez de la transparence, de la pédagogie sur la façon dont ils fonctionnent, sur le sens de leur choix. En effet, vos actions quotidiennes, sur le Web, sont dictées par des algorithmes : le livre recommandé sur Amazon, le lien publicitaire en haut de la liste de résultats Google, vos trajets en voiture. Ces micro-ajustements (anodins pris individuellement) assèchent progressivement vos libertés individuelles. *A minima*, il va vous falloir mieux comprendre les algorithmes, comment ils vous entourent et influencent vos actions. Lorsque vous naviguez sur le Web pour réserver un billet d'avion ou une chambre d'hôtel, vous ne voyez pas les

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

mêmes prix que les autres internautes, pour un lieu ou un moment de réservation pourtant identiques. C'est la mission de la Direction générale de la concurrence, de la consommation et de la répression des fraudes (DGCCRF) d'y veiller. Les prix sont-ils conçus à partir de critères véritablement objectifs ? Comme dans les premiers temps de la conquête de l'Ouest américain, le Web est encore un territoire peu régulé, et certains en profitent !

Un autre exemple, pour finir (nous l'espérons !) de vous convaincre de l'importance, pour le citoyen comme le consommateur, de comprendre les algorithmes. Prenez une application GPS qui permet de vous diriger à pied, en voiture ou en transports en commun. Elle vous propose un trajet. Comment s'assurer que cet itinéraire est bien celui qui optimise le temps de trajet, et pas celui qui vous fait passer par un maximum de points d'intérêts commerciaux ?

Copyright © 2018 Dunod.

Parce que l'algorithme engage votre liberté, vous devez interroger son éthique. Prenez les voitures autonomes : un célibataire, au volant de ce type de voitures, fait face, à grande vitesse, à une famille de cinq piétons. La voiture doit-elle continuer sa route et percuter ces piétons ? Ou bien se jeter dans le ravin de droite ? Qui décide du suicide ou du meurtre ? Le codeur de l'algorithme de conduite ? Le conducteur, au moment de l'achat de la voiture ? Si vous affectionnez ces dilemmes moraux, induits par la machine, précipitez-vous sur le jeu (gratuit et libre d'accès !) *Moral Machine*[1].

ILS L'ONT FAIT !

Une fois n'est pas coutume, le secteur public a pris les devants sur le sujet de la transparence des algorithmes.

En France, plus d'un million de collégiens et de lycéens participent chaque année aux procédures d'affectation automatisées Affelnet, pour les admissions post-Bac. Derrière ces affectations, de nombreux algorithmes, qui prennent en considération quelques variables majeures : résultats scolaires, proximité géographique, statut de boursier… sans qu'elles ne soient explicites. Parfois, l'opacité est difficile à accepter. D'autant plus lorsque le moteur se grippe, et que de nombreux étudiants restent sur le carreau, comme à la rentrée 2017.

La Loi pour une République Numérique, adoptée en septembre 2016, a fait un pas dans le bon sens (celui de la transparence). L'article 4 crée un droit pour le citoyen de connaître les règles de mise en œuvre de tout algorithme utilisé pour une décision administrative le concernant.

En conclusion : la data est un enjeu transformationnel. Posez-vous la question de la stratégie à adopter, du cap collectif à donner à vos équipes. Vous pouvez choisir de faire de la data un défi opérationnel,

1. http://moralmachine.mit.edu/hl/fr

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

une nouvelle opportunité pour optimiser vos procédures. Vous pouvez également, en élargissant l'angle, considérer que la data est une façon de faire autrement votre métier, d'ouvrir une nouvelle ligne de revenu pour votre organisation. Quelle que soit votre position, le mouvement n'est pas une option. Les attentes du régulateur, les attaques des *hackers*, vous obligent à réagir. Sur ce sujet, l'avenir est écrit. Préférez donc l'anticiper, plutôt que de le subir. Une réponse vaut mieux qu'une réaction !

Copyright © 2018 Dunod.

2 CHAPITRE Bâtir une culture de la data

Fixer la stratégie est une chose. Réussir son exécution en est une autre, et autrement plus complexe. Si votre entreprise a la data et son partage dans ses veines depuis sa fondation, alors passez ce chapitre. Si ce n'est pas le cas, alors le facteur clé de réussite ou d'échec de votre stratégie data sera votre capacité à faire évoluer votre culture d'entreprise, et ce davantage que les problématiques d'investissements financiers ou humains. Une culture d'entreprise a cette étrange qualité que d'être à la fois la chose la plus partagée et la moins formalisée. Il ne suffit pas de changer les organigrammes pour changer les comportements. Votre défi sera de taille. D'autant que la data, vous le verrez, déclenche généralement deux réactions opposées : l'indifférence (*Ce n'est pas mon sujet.*) ou, au contraire, l'instinct de possession (*C'est ma data ! Pas touche !*). L'enjeu : évitez à la fois Charybde et Scylla. Et aidez votre entreprise à créer les conditions d'une culture qui favorise la collecte, l'utilisation et le partage des data.

Copyright © 2018 Dunod.

Copyright © 2018 Dunod.

Permettez-nous de vous avertir que, dans ce chapitre, nous ne traiterons pas exclusivement de *Big Data*, mais de data tout court. L'enjeu de la culture dépasse en effet le seul périmètre des algorithmes, des *Data Scientists* et autres *datalakes*, dont vous entendrez parler dans les prochains chapitres. C'est de la data en général dont il s'agit, à savoir toute information, sémantique comme chiffrée, formalisée et susceptible d'être partagée dans votre entreprise. Les liens que vos collaborateurs entretiennent avec les chiffres, la circulation plus ou moins fluide des informations entre eux ou encore l'habitude plus ou moins ancrée qu'ils ont de formaliser, sont autant de critères clés pour apprécier votre culture d'entreprise. En termes de data, ils décideront du succès ou de l'échec de votre stratégie *Big Data*.

1 | LA DATA, CE N'EST PAS UNE ARME !

EN VOILÀ UNE BONNE QUESTION ! Comment convaincre mes collaborateurs que les chiffres ne servent pas qu'à évaluer leur performance en fin d'année ?

Encore partout, l'information, c'est le pouvoir. C'est d'ailleurs sur ce principe que s'est largement construite l'autorité managériale. Pour creuser l'idée, lisez l'excellent ouvrage de François Dupuy, *La Faillite de la pensée managériale*[1]. Connaître les prochaines nominations ; apprendre avant les autres des éléments sur la stratégie ; en savoir davantage que ses collaborateurs sur les chiffres d'activité de l'entreprise… autant de démonstrations de pouvoir pour le manager. Cette idée peut tuer l'intelligence collective, et le succès de votre aventure data. La rétention d'information, c'est l'ennemi. La valeur d'une data tient précisément à sa dimension transverse, aux liens que vous allez faire avec d'autres data, sur d'autres territoires. *Que se passe-t-il lorsque nous lions les data de performance aux augmentations individuelles, en ressources humaines ?*

1. François Dupuy, *La Faillite de la pensée managériale. Lost in management 2*, Seuil, 2015.

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

Les data de ventes aux profils des clients qui achètent le produit ?

La météo à nos besoins en termes de stocks ?

D'où la nécessité de casser les silos, en rendant transparentes et accessibles les data au plus grand nombre. Reprenez les trois axes stratégiques détaillés dans le chapitre « Définir votre stratégie » (p. 1-30). Les data de l'une de vos équipes peuvent aider à la performance de l'équipe voisine, « pour mieux faire ». Et l'addition des gisements de data de toutes vos équipes vous permettra, avec un regard large et englobant, de comprendre comment « faire autrement ».

Enfin, une maîtrise transverse de la data, de sa complétude, de sa qualité et de ses finalités, est indispensable pour « ne pas disparaître ». Pour ces trois raisons, ouvrez les vannes ! Transparence toute !

Voici un vrai paradoxe, mais qui n'est pas si étonnant, dès que l'on touche à des éléments de culture. Les entreprises les plus structurées autour de l'évaluation chiffrée de la performance (les fameux KPI, pour *Key Performance Indicators*[1]), ne sont pas les mieux armées pour réussir ce changement de culture. Pensez aux organisations anglo-saxonnes, tout particulièrement. Pour les collaborateurs, la data y est alors automatiquement connotée évaluation, et donc sanction, positive comme négative. Autrement dit : une culture « de chiffres » n'est pas, et ne prépare pas spécialement à, une culture « data ».

ILS NE L'ONT PAS FAIT…

Usons, cette fois-ci, d'une référence audiovisuelle. La série américaine *The Wire*, diffusée dans les années 2000, a pour sujet la criminalité, sous toutes ses coutures, dans la ville de Baltimore. Vécue à travers ceux qui en font l'expérience : policiers, trafiquants, politiques, enseignants et journalistes… La troisième saison est consacrée à la police, et à son fonctionnement. Elle illustre parfaitement l'impasse d'une organisation, lorsqu'elle ne vit que

1. Indicateurs clé de performance.

Copyright © 2018 Dunod.

pour un unique indicateur de performance (KPI), ici le taux de criminalité. Pour le coup, les réunions de l'État-Major filmées dans la série sont loin d'être des *safe meeting*, mais plutôt des séances testostéronées, rudes, hiérarchiques, dans lesquelles le KPI est l'alpha et l'oméga de la sanction. Plus grave : l'obsession chiffrée amène les policiers à délaisser le vrai sujet, à savoir la sécurité de leurs concitoyens. En se préoccupant uniquement des leviers (la plupart artificiels !) pour améliorer le KPI, ils en oublient la réalité. Le fait de ne plus déclarer un grand nombre de petits délits, pour faire baisser le taux, ça n'améliore pas la sécurité publique ! La vérité d'un KPI, c'est aussi ce qu'il n'est pas. Pour creuser les effets pervers d'une culture d'évaluation, nous recommandons l'excellent ouvrage collectif *Statactivisme ou Comment lutter avec des nombres*.[1]

Pour mener à bien cette transformation de la posture managériale, commencez par votre *top management*. Mais comment faire pour bouger rapidement les lignes ? Créez des premières instances, au sein de ce *top management*, dans lesquelles les data des uns et des autres sont mises sur la table, discutées, challengées, en sécurité. Nous parlerons (parce que la pratique nous vient de l'autre côté de l'Atlantique), de *safe meeting*. Donnez l'exemple, pour démarrer. Présentez les chiffres d'activité de l'entreprise, et parlez-en différemment.

Pourquoi ça monte ? Pourquoi ça baisse ?

Qu'est-ce que cet indicateur nous dit vraiment ?

Ne faudrait-il pas le lier à une autre data, pour lui donner son sens ?

Quels indicateurs, importants, ne sont pas sur la copie ?

Favorisez l'interaction à chaud. En faisant intervenir, par exemple, un de vos *Data Scientists*. En lui permettant de jouer dynamiquement avec un lot de data pour confirmer ou infirmer des premières pistes d'investigation discutées autour de la table. L'objectif de ces *safe meeting* est double.

1. *Statactivisme ou Comment lutter avec des nombres*, Zones, 2014.

© Dunod - Toute reproduction non autorisée est un délit.
Copyright © 2018 Dunod.

D'abord, montrer qu'un échange data peut échapper à la seule et réductrice question de l'évaluation. Puis, éduquer le collectif à tirer du sens, tout simplement, des data.

Au fil du temps, vous pourrez élargir ce type de réunions dans tout le reste de votre entreprise. À chaque équipe le soin d'animer ces échanges libres. Un conseil : invitez fréquemment des personnes externes (ou d'équipes voisines) à donner leur avis, à challenger, avec leur regard frais.

COMMENT ÇA FONCTIONNE, CONCRÈTEMENT ?

Quelques autres conseils pour réussir ces *safe meeting* : délaissez les nombres absolus, préférez-leur les tendances.

Favorisez les ordres du jour libres.

Faites participer des collaborateurs externes, non directement concernés.

Allez le plus rapidement possible de la data à l'action, en usant de la règle du « Et alors ? » : si, au bout de trois « Et alors ? » successifs, vous n'aboutissez pas à une recommandation concrète d'action, alors l'indicateur n'est pas utile.

Essayez cette règle, vous l'adopterez ![1]

La data n'est pas une arme. Veillez donc à l'afficher de façon transparente, le plus souvent et le plus largement possible. Sur les écrans de communication interne, dans les couloirs, au côté du cours de l'action ou de la citation du jour, sur les écrans de veille de vos collaborateurs. Pour ce faire, évitez les tableaux de chiffres austères, voire menaçants, dont tout le monde a trop l'habitude. Osez les couleurs, les formes… et usez, usez, de la *Dataviz* !

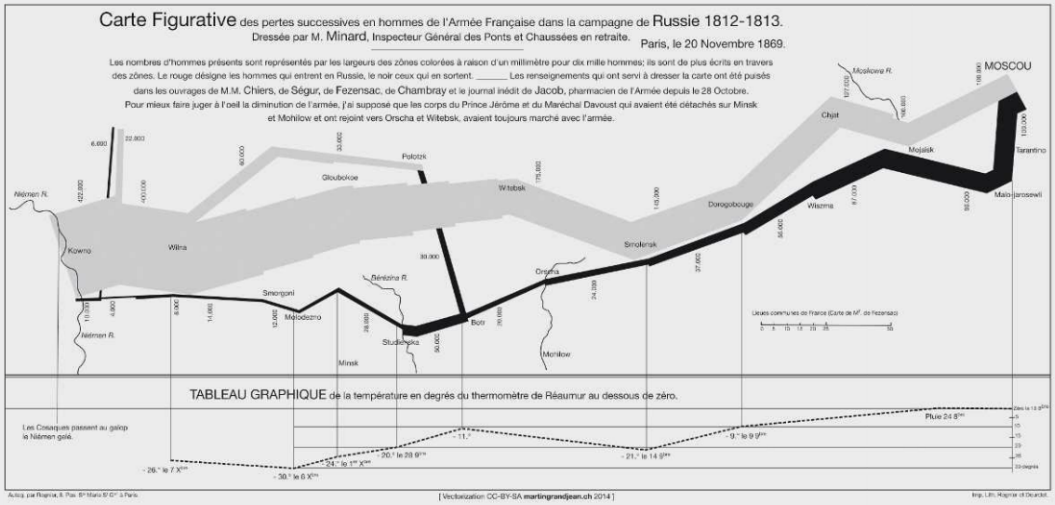
1. Kaushik Avinash, « Kill Useless Web Metrics: Apply The "Three Layers Of So What" Test ».

Copyright © 2018 Dunod.

QUEL AFFREUX JARGON !

DATAVIZ

La visualisation des data n'est pas qu'une histoire de coloriage. Vous faites face, quotidiennement, à un déluge d'informations, de *reporting* ou d'indicateurs, contribuant à ce que certains qualifient d'« infobésité ». Dans ce contexte, ne rajoutez pas du désordre. Donnez du sens, de la forme.
La *Dataviz* vous permet, en un clin d'œil, d'extérioriser les éléments intéressants d'un tableau de data, une anomalie, un comportement récurrent… d'où l'importance des *designers* dans une équipe dédiée à la data (rendez-vous dans la partie « La diversité, avant tout ! », p. 85).
Un exemple ? Comment toucher du doigt, en un clin d'œil, la débandade napoléonienne lors de la campagne de Russie, de 1812-1813 ? La réponse en image, sans aucune ambiguïté[1] :



Pour exprimer plus rapidement le sens, n'ayez pas uniquement recours à la forme visuelle. La mise en musique permet aussi d'identifier des anomalies. C'est ce qui s'appelle la « sonification », à savoir la représentation et l'émission de data sous forme de signaux acoustiques non verbaux, à des fins de transmission ou de perception d'information. L'ouïe est un sens

■■■

1. https://en.wikipedia.org/wiki/Data_visualization#/media/File:Minard%27s_Map_(vectorized).svg

© Dunod - Toute reproduction non autorisée est un délit.
Copyright © 2018 Dunod.

qui a ses avantages. Elle permet plus de précision et de couverture que la vue.

Prenez l'exemple simple d'une anomalie sur une série de mesures. Visuellement, vous pourriez illustrer la série par une succession de points noirs, séparés à intervalles réguliers, et l'anomalie par une réduction ponctuelle de la distance de l'intervalle, entre deux points noirs. Parfait, mais au bout de plusieurs centaines de points noirs qui défilent sous vos yeux, votre vigilance va décroître.

Passez au son : votre oreille sera plus attentive, et plus longtemps, au rythme, et donc à son accélération, à un point donné (celui de l'anomalie) entre deux notes.

Autre exemple : la hausse ou la baisse d'un indicateur chiffré, mise en son, pourra se traduire par une gradation plus ou moins forte du volume…[1]

IL TÉMOIGNE !

MARTIN DANIEL, DATA SCIENTIST@AIRBNB

C'est avec la logique que nous prouvons et avec l'intuition que nous trouvons. Cette phrase du mathématicien Henri Poincaré résume bien l'enjeu de la diffusion de l'information au sein des entreprises : donner accès aux données, c'est permettre à chacun de prouver son intuition et ainsi de faire avancer l'innovation.

Les exemples sont légions dans les entreprises de la Silicon Valley. Un cas fameux est celui de la création de Gmail. Dans une vidéo de 2008[2], l'un de ses créateurs, Paul Buchheit, montre comment le recours constant à l'avis de ses premiers utilisateurs l'a amené à développer un service désormais utilisé par plus d'un milliard d'individus[3].

1. « Scientific data might be filled with important things waiting to be discovered », *The Economist*, 19/03/2016.
2. Paul Buchheit at Startup School 08 – https://www.youtube.com/watch?v=EZxP0i9ah8E
3. « Gmail Now Has More Than 1 Billion Users », Techcrunch.

Copyright © 2018 Dunod.

Chez Airbnb, cette conviction s'est traduite par la constitution d'une infrastructure d'hébergement de données robuste et fiable, mais aussi accessible à tous[1]. De nombreux outils ont été développés en ce sens : Superset offre un environnement de requête SQL[2] et des outils de visualisation permettant à tout employé d'accéder à l'ensemble des tables et des bases de donnvées de l'entreprise, le service Dataportal[3] indexe et offre une fonctionnalité de recherche permettant de naviguer sur l'océan des bases de données, connexions ou résultats d'analyse créés par ses pairs. Enfin, la plateforme d'expérimentation ERF[4] permet de lancer et de mesurer simplement le résultat de tests A/B.

Ces outils s'accompagnent d'un important effort de formation – aussi appelé *Data University*[5] – qui permet là encore à tout employé qui le désire de se former sur des sujets allant des rudiments du langage SQL aux techniques avancées de l'apprentissage automatique, en passant par la maîtrise des concepts statistiques de base.

Aujourd'hui, près d'un tiers des collaborateurs d'Airbnb dans le monde effectuent au moins une fois par semaine une requête sur la plateforme et ont ainsi les moyens de prouver leur intuition.

Pour finir de convaincre vos collaborateurs, portez haut la méthode de l'A/B *testing*. Qu'est-ce à dire ? Chaque décision *business* vous amène à faire un choix, donc à assumer une part de risque, entre une option A et une autre option B. La création d'une nouvelle campagne publicitaire, l'ergonomie de votre site Web, le mode de rémunération de vos conseillers en plateformes téléphoniques… *Via* un A/B *testing*, vous mettez en place l'option A pour un segment d'individus (clients ou

1. Superset, Airbnb
2. *Structured Query Language* – langage informatique servant à exploiter des bases de data structurées.
3. Data Infrastructure at Airbnb
4. Scaling Airbnb's Experimentation Framework
5. How Airbnb Democratizes Data Science with Data University

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

prospects) et en même temps l'option B pour les autres individus. Et vous comparez les résultats, après une courte période, sans biais statistique. Cette méthode nourrit et permet une culture du *test & learn*, donc n'hésitez pas !

COMMENT ÇA FONCTIONNE CONCRÈTEMENT ?

La méthode de l'A/B *testing* a connu une forte accélération avec le développement du digital. Plus que tout autre média, le Web se prête à l'expérimentation et au test. Il se rapproche à ce titre du monde du marketing direct, habitué à l'optimisation continue, par le biais de mécanismes toujours plus poussés de ciblage.

Mettez en concurrence plusieurs éléments de votre site (modules, visuels, pages, entonnoirs…) présentés respectivement, dans le même laps de temps, à différents segments d'internautes, choisis au hasard, et pour qui le test est indolore. À chaque visite sur votre site, l'outil de *webtesting*, en fonction du *cookie* de l'internaute (pour plus de précisions, voir la partie « Les data digitales, tu chériras ! » p. 115), prendra la décision de lui présenter une version A ou B.

La pratique du *webtesting* présente deux avantages majeurs, et d'abord un indiscutable pouvoir de démonstration. Les éléments comparés les uns aux autres le sont de façon simultanée et invisible pour vos internautes. Les risques de biais de saisonnalité ou d'usage sont donc écartés. Les intuitions de vos équipes, quant à la taille et au libellé que doivent prendre tels ou tels boutons d'action, ou encore la présentation optimale d'une offre en page d'accueil, s'effacent derrière la décision et le plébiscite des internautes. Enfin, le *webtesting* fait évoluer les mentalités, au sein des équipes. Pour les directions marketing, il s'agit d'accepter que les choix effectués pour un site Web soient challengés, contredits par les internautes. Pour les directions informatiques, d'accepter qu'un site ne reste pas figé, mais doit être multiple et agile. Le *webtesting* revient à mettre l'internaute au centre et à faire progressivement de la data l'impulsion et l'horizon de vos décisions marketing.

Copyright © 2018 Dunod.

2 | LA COLLECTE POUR TOUS !

EN VOILÀ UNE BONNE QUESTION ! Comment s'assurer que nous collectons bien toutes les data utiles à notre activité ?

Tous les jours, vos collaborateurs effectuent des activités dans l'entreprise. Mais posez-vous la question : *Quelle part de leur travail produit de la data ? De l'information ?* À y réfléchir, probablement très peu. Trop peu. Pourtant, vos collaborateurs produisent beaucoup de data, tirées de ce qu'ils font. Mais ces data restent informelles, captées, puis trop vite oubliées, par la mémoire de chacun. Le sourire d'un client ; la répétition d'un *bug* sur un système informatique ; le résultat d'un A/B *testing* sur un site Web… L'enjeu ? Formaliser ! Nous aurions pu rajouter : lorsque c'est utile.

Or, il est difficile pour un collaborateur, ou même une équipe, de juger de ce qui est utile au collectif. Même pour des secteurs *a priori* peu centrés sur le client final, le digital va quand même vous rapprocher de vos clients. Et plus la relation avec eux va devenir interactive, chaude… plus votre capacité à personnaliser leur expérience va faire la différence.

Or, pour personnaliser, il faut connaître, anticiper, s'ajuster… donc avoir collecté beaucoup de data ! Nous partageons, ici, une cible idéale. Le nerf de la guerre, en matière de collecte data, ce sont vos outils. S'ils rendent la collecte naturelle à vos collaborateurs, en l'intégrant dans leurs gestes-outil, c'est déjà gagné !

ILS L'ONT FAIT !

Dans les deux premières saisons de *Twin Peaks*, la série culte de David Lynch, le héros est un agent du FBI, Dale Cooper. Il ne se contente pas d'être brillant, toujours impeccablement vêtu, amateur de café noir et gendre idéal (en tout cas dans les deux premières saisons !) : il raconte systématiquement tout ce qu'il fait, tout ce qui lui arrive, à Diane, sa collaboratrice, *via* un dictaphone qui ne le quitte jamais. Belle illustration de formalisation !

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

QUEL AFFREUX JARGON !

MARKETING 1TO1

Pendant des décennies, pour comprendre leurs clients et anticiper leurs besoins, les directions marketing ont eu recours à des panels de consommateurs, avec une louable intention : écouter les clients.
Malheureusement, ces études quantitatives fonctionnent sur la base de l'échantillonnage. Le comportement observé sur un nombre réduit de clients, censé être représentatif, est ensuite généralisé à tous les clients. Impossible d'utiliser ces études pour en déduire une action personnalisée, sur un client identifié comme tel.
Ce qui a changé aujourd'hui : les technologies *Big Data* permettent désormais de traiter l'intégralité des data issues d'un portefeuille client, sans avoir à en passer par un échantillonnage. L'ensemble des interactions avec vos clients devient, de fait, l'objet d'une vaste étude. Avec, contrairement aux panels, une appréciation individuelle des profils et donc des actions marketing à entreprendre, sur des clients bien identifiés.
C'est ce qui est appelé le marketing 1to1.

Une illustration de collecte, concernant les plateformes téléphoniques, si vous en utilisez. N'attendez plus, enregistrez l'ensemble des discussions. Ayez recours à des outils d'enregistrement, permettant d'opérer du *speech to text*, à savoir une retranscription automatique des discussions orales en textes. Les usages *business* sont innombrables, notamment pour améliorer la satisfaction-client, en détectant mieux les irritants exprimés. Et n'oubliez pas que demain, ou après-demain, vous aurez très probablement à développer des *bots*, à savoir des robots capables d'interagir, *via* un Chat ou sur les réseaux sociaux, avec des clients de façon automatisée, en usant un langage naturel.

Ces *bots* doivent apprendre à comprendre les mots de vos clients pour répondre à leurs sollicitations. Il faut les éduquer. L'aliment qui va permettre d'entraîner le robot, ce sont justement toutes les discussions réelles, enre-

Copyright © 2018 Dunod.

gistrées, menées par vos gestionnaires en plateformes. D'un coup, ce qui apparaissait inutile et fastidieux (collecter systématiquement l'ensemble de vos discussions téléphoniques) devient l'une des clés de votre compétitivité, dans un contexte de rapide automatisation du marché.

En interne, présentez, expliquez et mettez en avant le concept-socle de collecte « pour tous ». Une exigence d'autant plus critique, si vous entretenez plusieurs canaux d'interaction avec vos clients : distribution physique, écrans digitaux et plateforme téléphonique. Dans un contexte omnicanal, voici l'ambition collective que vous pourriez, par exemple, vous fixer (la formulation est un peu administrative, mais toutes les notions importantes y sont présentes) : *Chaque point de contact avec nos clients, par nous ou par les partenaires de notre entreprise, donne lieu à une collecte cohérente, fluide, et de qualité, de toutes les data susceptibles d'enrichir la connaissance que nous avons d'eux… data structurées qui, diffusées dans nos systèmes et partagées avec pertinence entre les acteurs de la relation client, permettront, dans le respect de la réglementation, d'enrichir notre proposition de valeur, depuis l'acquisition jusqu'à la fidélisation de nos clients.*

Attention ! Devenir une entreprise « apprenante » n'est pas simple ! Pour être à la hauteur, mettez vos collaborateurs en capacité de collecter les data. Pour ce faire, voici deux conditions minimales à réunir.

METTRE EN PLACE DES OUTILS

Les outils doivent vous permettre de récolter deux types de data. D'abord, les data indispensables à votre activité et à la réglementation (pensez à l'exigence de connaissance du client). Mais aussi, les data « de côté », qui ne vous sont pas immédiatement indispensables et utiles, mais qui pourraient l'être demain. Vous appuyer sur un outil de CRM[1] unique vous facilitera grandement la tâche. Sur son poste de travail en plateforme téléphonique, sur son iPad lors d'une vente en porte-à-porte… chacun de vos collaborateurs doit, *a minima*, pouvoir signifier,

1. Customer Relationship Management – interface qui rassemble les data clients.

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

dans un champ de formulaire vide, par quelques phrases ou quelques mots, les informations les plus intéressantes.

Quelle part de nos collaborateurs peut, dans l'exercice de son métier, remonter une data intéressante, quelle que soit sa nature ? Plus vos formulaires de collecte seront sophistiqués, plus les data, une fois dans vos systèmes informatiques, seront facilement utilisables. Pensons à une série de questions simples, intuitives auxquelles vos collaborateurs pourraient rapidement répondre : *Client satisfait ? Oui ou non ? Déménagement en vue ? Oui ou non ? Suspicion de fraude ? Oui ou non ?* Le graal ? Que cette collecte ne nécessite pas un effort supplémentaire dans l'équipe, mais fasse partie intégrante du geste-outil lié à son activité. La parfaite illustration est la vente de produits d'épargne. Votre conseiller doit connaître son client, de façon à lui apporter de justes et honnêtes conseils. À cette fin, lors de l'entretien, il récolte de nombreuses data, *via* un module de type « *Know Your Customer* » (connaître son client) : salaires, situation du conjoint, patrimoine immobilier… Un exercice de collecte utile pour l'entreprise, et accepté comme tel par le client. C'est une configuration idéale.

MOTIVER VOS ÉQUIPES

Collecter davantage vous prendra du temps. Et le temps… c'est de l'argent ! Motivez vos collaborateurs en évoquant l'ambition collective, votre stratégie d'entreprise. Bien qu'indispensable, ce n'est probablement pas suffisant.

Pour garder des niveaux élevés, chaque année, de collecte de data (en quantité comme en qualité), appuyez votre programme sur une rémunération spécifique, et ajustez le partage de la valeur. Sans trop caricaturer, vous avez deux populations bien différentes à convaincre : l'interne et l'externe.

Pour la première, celle de vos collaborateurs, la rémunération reste un levier facile à activer, et à effet plutôt immédiat. Par exemple : sur vos plateformes téléphoniques, collecter plus de data prendra tant de secondes supplémentaires à chaque appel. Ajustez donc les objectifs

Copyright © 2018 Dunod.

de productivité en conséquence ! Généralement, vos collaborateurs internes n'ont rien contre le principe de mieux collecter et partager les data. Ils demandent simplement (et à juste titre) que leurs objectifs et rémunérations évoluent en cohérence.

Pour la seconde population, l'externe, la partie est moins facile. Il s'agit de vos partenaires, de vos prestataires, ou encore de vos distributeurs. Ils peuvent s'inquiéter de perdre la relation client s'ils partagent ainsi les data avec vous. D'où, parfois, la multiplication de comportements déviants : des numéros de téléphone clients qui restent sur des annuaires personnels, le recours à des messageries e-mail « à part »… Dans ces cas-là, le levier de la rémunération immédiate, au temps passé, ne suffit pas. Il vous faudra repenser la question, plus globale, du partage de la valeur.

> ***EN VOILÀ UNE BONNE QUESTION !*** Comment intéresser économiquement mon écosystème externe à la valeur globale d'un client ?

Pour libérer les énergies de chacun, ne prenez pas le sujet par le petit bout de la lorgnette !

Vous avez besoin, avant de vous lancer, d'avoir une idée plus précise du potentiel de collecte ? Demandez à vos équipes de faire preuve d'imagination. Et de lister toutes les data qu'elles pourraient collecter, si c'était leur travail à temps plein. Puis, de les catégoriser ainsi : collectées et utiles ; collectées et inutiles ; non collectées et utiles ; non collectées mais inutiles… vous verrez des éléments intéressants, et ce dans les quatre catégories (même la dernière !).

La cartographie des data disponibles, en interne comme en externe, laissera transparaître, à coup sûr, beaucoup de zones grises, à savoir des gisements de data précieux, sur lesquels la question de la propriété ou de l'accessibilité se pose. Cet exercice est justement l'opportunité de trancher. *Parle-t-on de data critiques pour mon entreprise ? Ou d'importance plus marginale ?* En fonction des réponses apportées, déclenchez un mouvement de ré-internalisation de certaines data. Prenons

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

un exemple tiré de l'univers marketing. Si vous investissez fortement en publicité, vous êtes sûrement accompagné par une agence d'achat média. Or, la publicité, notamment digitale, génère beaucoup de data : impressions de bannières, clics, visites, temps de diffusion… des gisements essentiels dans la compréhension du cycle d'achat et le profilage de vos clients et prospects. Eh bien, posez-vous deux questions : *Où sont ces data ? Pouvons-nous y avoir facilement accès, au niveau le plus fin que nous souhaitons ?* Ces data sont collectées par les outils digitaux d'achat média, et parions que vous devrez passer, vous ou vos équipes, par votre agence pour y accéder…, avec le risque de friction qui en découle. Si c'est le cas, un conseil : ré-internalisez toutes les data publicitaires, en contractualisant directement avec les éditeurs d'outils digitaux.

ILS L'ONT FAIT !

La collecte des data a toujours un coût, plus ou moins caché. Au nom de l'ambition fixée, il faut assumer. Pensez aux programmes de fidélité, dans la grande distribution.

Walmart a récemment estimé le coût de son programme de fidélité à plus d'un milliard de dollars, principalement imputé à tous les avantages et réductions concédés. Voilà pour le coût.

Pour le gain : ce programme a permis au distributeur américain (entre autres) d'accélérer fortement ces dernières années la compréhension des paniers d'achat… et d'affiner, ainsi, la performance de ses algorithmes de recommandation, qui foisonnent aujourd'hui dans son écosystème digital !

Autre secteur qui bouge, celui du transport de marchandises. Certains transporteurs s'en sont aperçus : leur métier ne se cantonne pas à transporter des produits. Ils ont à portée de leurs mains, dans leurs camions, leurs trains et leurs avions, un gisement incroyablement riche de data sur les marchandises transportées. Ils ont donc commencé à monétiser ces data, en les vendant aux départements logistiques et *business intelligence* de leurs clients

Copyright © 2018 Dunod.

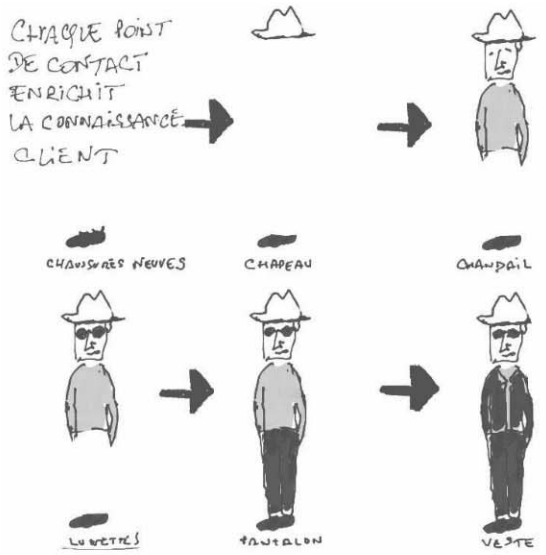
QUEL AFFREUX JARGON !

TYPER !

Les technologies *Big Data* vous permettent de collecter des data non structurées. C'est déjà une avancée majeure mais, ne vous trompez pas, plus vos data sont structurées, mieux vous vous porterez.

En aval, dans l'analyse, les statisticiens passent généralement trop de leur temps à nettoyer et interpréter les data. Autant de temps qui n'est pas utilisé à optimiser la performance du modèle.

Pour leur faciliter la tâche, il est idéal que chaque data collectée soit typée, et ceci correctement. Autrement dit, qu'il existe une méta information indiquant explicitement la catégorie de la data collectée. Une méta information, c'est l'équivalent du titre de la colonne, pour reprendre l'image d'un tableau Excel. *La data «26/09/85»*, c'est une *«date de naissance». La data «suivi de dossier»*, c'est un *«motif d'appel». La data «divorce»*, c'est un *«moment de vie»*… Dans la mesure du possible, vos collaborateurs doivent donc collecter et typer la data dans le même temps. Projetons-nous, demain, sur l'intelligence artificielle. Une machine apprendra plus facilement sur des data structurées, donc typées. Le temps de typage, à chaque collecte, n'est donc pas un temps perdu !



Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

3 | LA DATA, UN PATRIMOINE VIVANT !

Peu de traces, dans votre rapport d'activité annuel, de la data ? De sa qualité, de sa complétude, de votre niveau élevé de conformité à la réglementation ? Ça va changer ! Afficher le nombre de ses clients, c'est bien. Montrer à quel point vous les connaissez, c'est encore mieux. La data devient un actif à part entière, à valoriser comme tel, auprès des parties prenantes de votre entreprises (collaborateurs, régulateurs, propriétaires et autres actionnaires…).

Une fois collectées, les data doivent être traitées, échangées, utilisées par vos équipes, comme un bien précieux, car elles forment un patrimoine commun. C'est l'un des paradoxes actuels, dans nombre d'entreprises : la data est omniprésente, mais souvent orpheline. Posez des questions simples autour de vous, et jugez-en : *Qui est responsable de la qualité de l'adresse postale de nos clients ? Un client veut récupérer toutes les data que vous avez sur lui, qui s'occupe de les lui restituer ? Qui a le droit de changer l'adresse e-mail d'un client ? Qui peut, sous quelques minutes, dire dans quels systèmes informatiques résident ces data ?*

Historiquement, les data sont organisées en silos, même dans des organisations de petite taille. Vous trouvez les data commerciales d'un côté, les data liées à la fabrication de vos produits de l'autre, les data administratives à côté, les data relatives aux ressources humaines à l'autre bout… Les rassembler, créer des ponts entre elles, c'est bien l'enjeu (détaillé dans le chapitre « Libérer vos énergies » p. 79-96). Pour créer de la responsabilité collective sur le sujet data, c'est à la Direction Générale de s'y atteler. Quelques arguments pour convaincre :

- **Pas le choix, sinon le régulateur va nous taper sur les doigts.** Efficace, et véridique (pour s'en convaincre, allez à la partie « De la réglementation, tu te préoccuperas ! », p. 103). Mais reconnaissez que c'est là une approche assez négative, vous n'allez pas susciter l'enthousiasme de vos équipes !

Copyright © 2018 Dunod.

- **1+1 = 3.**
  C'est l'approche didactique, micro. Illustrez, par quelques exemples concrets (et bien choisis !), que deux data séparées l'une de l'autre, une fois rassemblées, génère de la valeur. Déjà plus convaincant, même s'il manque la vue d'ensemble !
- **C'est notre stratégie d'entreprise.**
  Vous venez de faire le travail de lier la data à votre stratégie globale (les clés dans le chapitre « Définir votre stratégie » p. 1-30). Profitez-en, communiquez largement ! Cette vision vous permet de donner du sens, et d'embarquer. *Soucions-nous de nos data, pour que nous devenions ce que nous souhaitons devenir !*

Votre objectif : aucune data, dans vos murs, ne doit rester orpheline, c'est-à-dire sans une équipe ou un collaborateur responsable de sa complétude, de sa qualité et de son usage. Cette responsabilité se décline en objectifs explicites, de préférence chiffrés. Notion importante : le bout-en-bout. La responsabilité doit couvrir l'ensemble du parcours de la data chez vous, de sa collecte jusqu'à son usage. Par cette exigence, vous obligez ainsi vos équipes à une véritable transversalité, qui impactera d'ailleurs l'organigramme en place.

Point d'attention : la Direction des Systèmes d'Information (DSI) ne peut pas prendre cette responsabilité, même s'il est tentant de la lui donner, faute de mieux (rendez-vous à la partie « Bousculer, sans braquer ! », p. 81). Nous n'insisterons jamais assez : c'est au *business* de porter la responsabilité des data. C'est avant tout parce que vous souhaitez lancer cette action de communication, développer ce produit, faire évoluer cette chaîne logistique, que vous avez besoin de cette data, sous cette forme et à cette fréquence. Placez les équipes *business* sur le siège du conducteur, en transformant leurs objectifs. C'est vous assurer d'un changement culturel rapide.

La responsabilisation est une condition nécessaire de réussite, mais pas suffisante. Une gouvernance doit suivre, dans le temps. À vrai dire, les objectifs d'une gouvernance data sont relativement communs. En voici énoncés quelques-uns :

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

- **Fixez (et tenez dans le temps) vos objectifs de collecte.**
  Et ceci à la fois en complétude et en qualité. Responsabilisez les équipes et les collaborateurs sur le sujet. Attribuez des objectifs, suivez-les dans l'année, ainsi que le plan d'action associé. *Ce mois-ci, nous subissons une chute du nombre de moments de vie collectés par les plateformes de gestion ? Que se passe-t-il ? Bug d'outil, relâchement du management sur les objectifs, ou évolution des usages de nos clients ?*
- **Respectez la réglementation en vigueur !**
  En fonction de la taille de votre entreprise, vous ajusterez les ressources à dédier au sujet. Mais la contrainte est la même : respecter les principes, notamment ceux portés par le GDPR, prochain règlement européen (plus de détails dans la partie « La data pour ne pas disparaître », p. 23). Et nommer, parmi vos collaborateurs, un *Data Protection Officer* (DPO) (p. 53) !
- **Concevez des principes.**
  Acceptez de créer un peu de bureaucratie. Dans la gestion de ses data, votre entreprise a besoin de principes. Sinon, vous risquez du désordre, d'autant plus élevé sur un nouveau territoire comme celui de la data, peu connu des équipes en place. Imaginez le résultat, après plusieurs mois, chaque équipe traitant les data partagées à sa façon. Et c'est un risque très concret.
  Une illustration, avec la CSP (Catégorie Socio-Professionnelle). Deux de vos collaborateurs la collectent, chacun de leur côté, en utilisant des référentiels différents (l'un comportant des abréviations, l'autre pas). Soyez assurés qu'une fois dans vos systèmes informatiques, les « Agriculteurs exploitants » et « Agriculteurs expl. » compteront pour deux lignes différentes. C'est l'un des premiers principes, celui de cohérence dans la collecte. Voici des questions à vous poser, car elles nécessiteront très probablement d'édicter des principes pour y répondre : *Une data client fraîchement collectée par un collaborateur peut-elle automatiquement remplacer une plus ancienne (potentiellement collectée par un autre collaborateur) ? Quelles data sont partageables, et avec qui ? Peut-on commercialiser les data ?* Visez de piloter

Copyright © 2018 Dunod.

la gouvernance data de votre entreprise avec une dizaine de principes forts, pas plus.

- **Faites respecter ces principes.**
  Les principes posés, il faut s'assurer de leur bonne exécution. Dans chaque projet, vos équipes sont amenées à faire des choix opérationnels. *Quel fichier externe utiliser ? Quel format pour l'adresse postale ? Quel consentement du client, et sous quelle forme le récolter ? La collecte de l'e-mail client est-elle obligatoire ?* Ces choix sont éclairés par les grands principes, mais ne s'y réduisent pas. Ils procèdent par interprétation, et font très souvent des entorses. Votre gouvernance data doit-elle tout contrôler ? Non ! Elle serait bien incapable de mener à bien sa tâche. Elle ne peut pas tout voir. Vous devez vous organiser pour ne traiter que les choix opérationnels qui posent difficulté, pour lesquels un accommodement aux principes est souhaitable. Vous vous rendrez compte qu'un de vos principaux défis n'est pas tant de décider, que d'intercepter tous les projets qui nécessitent votre aide. Et de sensibiliser vos collaborateurs pour qu'ils aient le réflexe de remonter d'eux-mêmes les dilemmes.

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

COMMENT ÇA FONCTIONNE, CONCRÈTEMENT ?

Certains de vos principes, c'est le régulateur (GDRP) qui se chargera de vous les inspirer. Prenez le « droit à l'oubli », à savoir l'interdiction de garder dans vos systèmes informatiques certaines data client plus que nécessaire à l'exercice de votre proposition de valeur. Par exemple, les data liées à vos prospects (délai fixé à trois ans après une première sollicitation commerciale). Votre gouvernance data formalisera le principe global, et validera ses déclinaisons opérationnelles : les règles de suppression des data adaptées à chaque système informatique…

> **Chaque révolution s'évapore en laissant seulement derrière elle le dépôt d'une nouvelle bureaucratie.**
>
> **Franz Kafka**

Facteur clé de succès de la gouvernance : la bonne coordination entre trois de vos équipes. Le *business*, la Direction des Systèmes d'Information (DSI), et le réglementaire. Si ce n'est pas le cas, partez du principe qu'il y a un risque élevé de doublons dans le travail et dans les responsabilités.

Pour bousculer la culture, la formation reste un levier majeur. N'hésitez à vous appuyer sur la richesse des formations proposées en ligne (les fameux MOOC[1]). À quoi doit-on former les collaborateurs, au bout du compte ? La capacité à « trier » les data avec discernement est particulièrement critique. Qu'est-ce à dire ? Qu'à chaque étape de traitement de la data, des décisions sont prises par vos collaborateurs, en réponse à une question principale et qui ne varie pas : *A-t-on vraiment besoin de cette data ?* Or, une data, si elle n'est pas directement utile à l'équipe qui la collecte, le sera très probablement à l'équipe voisine plus tard. Pour créer les bons réflexes de tri dans votre organisation, il faut montrer à tous vos collaborateurs la vision d'ensemble. Que chacun touche du

1. *Massive Open Online Courses* – à noter, l'éditeur américain Coursera est particulièrement pointu sur les enseignements autour de la data.

Copyright © 2018 Dunod.

doigt la multiplicité des usages possibles, pour une data donnée. Mais ce n'est pas simple !

QUELS AFFREUX JARGONS !

DATA STEWART/DPO/DMO

Notions clé, que vous trouverez dans tout bon manuel de gouvernance data.

Les *Data Stewarts* sont des collaborateurs en ligne directe avec le *business* (dans le jargon : de « niveau 1 ») et les activités opérationnelles. Ils sont très proches des sources de collecte de la data (distribution, plateformes téléphoniques, écrans digitaux…). Leur rôle ? Être responsables des objectifs de collecte, tant en quantité qu'en qualité. Pour que cela fonctionne, les *Data Stewarts* doivent consacrer *a minima* un bon tiers de leur temps à cette mission data, et être valorisés en interne pour cela.

Le DPO, ou *Data Protection Officer,* est un collaborateur chargé du contrôle à un « niveau 2 ». La réglementation, et plus spécifiquement le GDPR, lui attribue un rôle très régalien. Il est probable que le DPO tienne également lieu de CIL, à savoir votre « Correspondant Informatique et Libertés », auprès de la CNIL. Le DPO est responsable de la bonne conformité de l'ensemble des opérations liées à la data, dans votre entreprise.

Le DMO ou *Data Management Office* est le nom communément donné à votre petite équipe, en central, qui s'assure que votre gouvernance data tourne, d'un point de vue opérationnel. Entre autres, c'est lui qui construit et diffuse les *reporting* de data (dont la collecte), et veille au bon respect des principes data, dans la multitude de vos projets en cours.

La remise en cause des silos, dans votre entreprise, implique également, dans une moindre mesure, celle des « experts », ceux qui fondent

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

leurs actions *business* sur l'intuition issue de leurs expériences en profondeur, peu transverses. Nous parlons, ici, de métiers aussi divers que celui de rédacteur en chef, de libraire, de sélectionneur de football ou de médecin… Ces métiers ne vont pas disparaître, ce n'est pas le propos. Pensez plutôt que ces experts vont devoir, dans l'exercice de leur métier, composer avec de la data dont ils n'ont pas l'habitude, et laisser une place à l'automatisation des recommandations, des suggestions…

Enfin, un dernier élément doit bousculer votre culture d'entreprise : l'esprit d'ouverture. Ouverture aux data externes, dans un sens comme dans l'autre. Historiquement, vos systèmes informatiques sont fermés. Seules les interfaces « maison » (postes de travail, sites Web, applications mobiles…) consommaient, à distance, vos data « maison ». La seconde vague de digitalisation, déjà là, vous force à exporter vos data sur d'autres écosystèmes, non propriétaires. Pensez au développement massif des services client sur les réseaux sociaux : Messenger, WhatsApp… les clients y consultent à présent leurs comptes bancaires, y achètent des billets de train, y formulent leurs réclamations. Vos data circulent, s'exportent et se lient sur d'autres carrefours d'audience, sur d'autres interfaces digitales que les vôtres. L'entre-soi, ça n'est plus possible.

UN EXEMPLE

La publicité digitale vous permet de cibler, avec toujours plus de précision, vos clients. Le principe est simple : vous transmettez les adresses e-mails de vos clients à Facebook et/ou Google. Ces derniers font le lien avec leurs propres utilisateurs (dont ils ont collecté les adresses e-mails *via* leurs comptes Facebook et Gmail), et vous suggèrent de les cibler avec vos messages publicitaires sur leurs sites. Levier de ciblage très puissant, mais qui vous oblige à transmettre, de façon massive et transparente, vos adresses e-mails clients en externe, et pas chez n'importe qui. Cette transmission brusque les habitudes, notamment sur la sécurisation des data. Et devrait faire, typiquement, l'objet d'une instruction par votre gouvernance data, évoquée précédemment.

Copyright © 2018 Dunod.

Avant de conclure, arrêtez-vous sur le concept, souvent entendu, d'entreprise « élargie ». Il a beaucoup de sens pour le sujet qui vous occupe, la data. Il dépasse les frontières, trop étroites, de votre bilan ou de votre structure juridique, pour recouvrir l'ensemble des data en interne comme en externe, qui vous sont accessibles. Imaginez la nouvelle dimension, à cette aune, que prennent tout d'un coup les grands acteurs digitaux, qui maîtrisent les principaux carrefours d'audience Web : Google, Amazon, Facebook… ils voient tout passer dans leurs tuyaux. Et vous, quelles data supplémentaires gagnez-vous à penser votre entreprise comme « élargie » ?

Mais attention, en vous lançant dans l'aventure *Big Data*, tout feu tout flamme, sans vous assurer des fondamentaux, vous prenez le risque du *Big Flop*, de l'aventure solitaire et de courte durée…

> **Tout seul, on va plus vite. Ensemble, on va plus loin.**
>
> **Proverbe africain**

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

Copyright © 2018 Dunod.

3 CHAPITRE Concevoir votre trajectoire

Vous avez défini une stratégie claire et installé les premières briques d'une culture d'entreprise favorable à l'usage, par tous, de la data. À présent, il vous faut une trajectoire, dessiner un chemin collectif. Une trajectoire n'a de sens que si elle vous est spécifique. Votre succès tiendra à l'alchimie entre votre ambition, votre environnement et l'état d'esprit collectif. Méfiez-vous des bonnes pratiques génériques.

Copyright © 2018 Dunod.

Copyright © 2018 Dunod.

Il est important de savoir dans quelles situations il faut éviter de se retrouver après quelques mois. En l'espèce, voici les Charybde et Scylla de la data :

- **Une foultitude d'expérimentations, sans suite.** La data porte une force disruptive pour votre entreprise, ses procédures et son modèle économique. Présentez-la comme telle à tous vos collaborateurs. Car sinon, vous aurez de grandes chances que le corps social cantonne l'aventure à quelques malheureux et isolés POC (*Proofs of Concept – démonstrations de faisabilité*), sans impact tangible sur le quotidien de vos collaborateurs et de vos clients. Bien entendu, les expérimentations sont utiles, notamment sur le territoire de la data, sur lequel les gisements de gains *business* ne sont pas encore bien identifiés. Mais dès le début d'une expérimentation, pensez industrialisation. Un principe simple à tenir : avant de lancer une initiative, assurez-vous bien, auprès du sponsor, qu'en cas de succès de l'expérimentation (dont les KPI sont à définir *a priori*), l'industrialisation du cas d'usage est bien prévue, et le budget en conséquence.
- **Un unique projet, mais quinquennal.** C'est l'extrême opposé à éviter, le grand projet unique, qui mobilise toutes les ressources, avec des perspectives de gains *business* à plusieurs années. En termes de data, comme pour les autres projets informatiques, craignez l'effet tunnel. Surtout si votre entreprise peut se le permettre, ne mettez pas tous vos œufs dans le même panier. Poursuivez plusieurs lièvres à la fois !

Une trajectoire data réussie est celle qui laisse une place à l'expérimentation « à froid », peut en se fixant systématiquement, en cas de succès, l'ambition d'une industrialisation future.

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

QUEL AFFREUX JARGON !

À FROID

Désigne une expérimentation data qui n'est pas directement connectée aux systèmes informatiques. Pour ce faire, il est fréquent d'exporter des data des systèmes, sous la forme de fichiers plats, puis de les analyser dans un environnement fermé, non lié à l'activité *business*. Vos *Data Scientists* et autres statisticiens peuvent alors jouer, triturer et arranger les data comme bon leur semble, sans impact réel. Le « à froid » est une étape quasi indispensable pour estimer les gains *business* à attendre d'une initiative data, avant son industrialisation. *A contrario*, plus une initiative data se « réchauffe », plus elle communique (parfois jusqu'en temps réel) avec vos systèmes opérationnels.

Copyright © 2018 Dunod.

1 | LE GRAND SOIR TECHNOLOGIQUE, C'EST CE SOIR !

> ***EN VOILÀ UNE BONNE QUESTION !*** Vous êtes bien sympathiques avec tous vos beaux outils *Big Data*… mais notre système informatique actuel, c'est un plat de spaghettis. Qu'est-ce que nous en faisons ?

Plus votre entreprise a vécu, plus lourd est son héritage informatique (votre *legacy*). Les mauvaises langues parleront même de dette informatique pour le désigner ! Rassurez-vous : les entreprises les plus récentes, comme Google et Facebook, doivent composer avec la même problématique de rigidité, voire d'obsolescence, de leurs systèmes informatiques passés.

La raison est simple : un système est bâti pour répondre aux besoins du moment. Aussi modulaire soit-il, il ne peut anticiper la diversité de vos besoins futurs. Ici, ne vous trompez pas de combat : l'enjeu n'est pas de faire « sans » votre *legacy*, mais bien de savoir comment faire « avec », au mieux. Précision importante, et structurante : la première option, celle de la table rase, à quelques exceptions près, n'est pas réaliste. Rebâtir un système informatique complet, c'est accepter de mettre votre activité en pause, ou au ralenti de longs mois.

Votre trajectoire data, à trois ans, fera apparaître différents cas d'usage, proposés par vos équipes et séquencés dans le temps. Cette approche itérative, en séquence, est la plus appropriée pour prioriser des initiatives *business*. Pour autant, elle touche ses limites en termes d'informatique. Il y a de grandes chances que le degré élevé d'éclatement et de rigidité de vos systèmes informatiques vous oblige, dans un premier temps, à un Grand Soir, à savoir à un investissement financier initial, plutôt conséquent, pour bâtir les fondations technologiques du *Big Data*.

C'est vrai, les gros tickets d'investissement informatique, de nos jours, n'ont pas la cote. Vous leur préférez (et vous avez raison) les enve-

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

loppes plus petites, sujettes à un pilotage plus agile par vos équipes de proximité.

Pour autant, si vous voulez ce démarrage *Big Data* rapide, vous devrez faire une exception à la règle. Pour vous en convaincre, relevons deux caractéristiques des cas d'usage *Big Data*, qui les rendront difficiles (voire impossibles) à intégrer dans votre environnement informatique classique.

DES DATA DIFFÉRENTES, AU MÊME ENDROIT

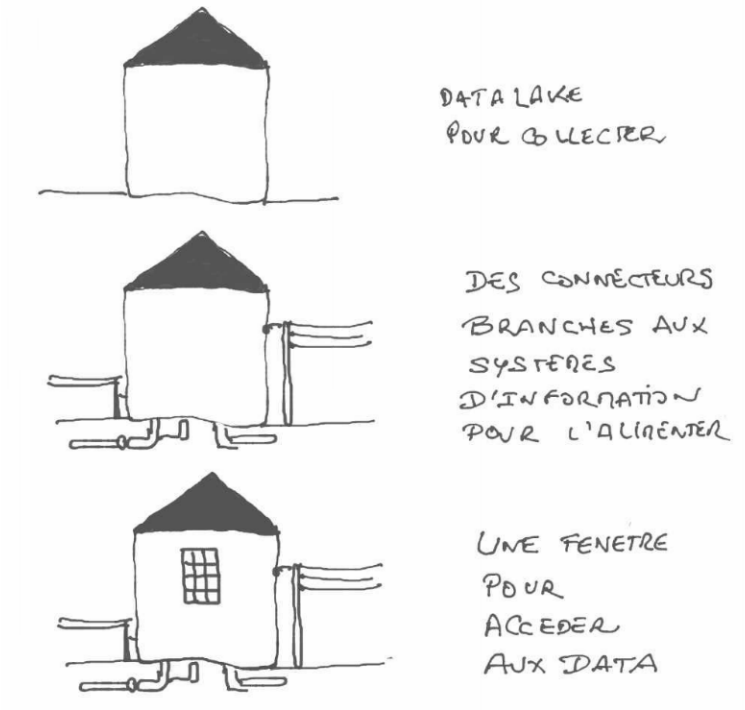
Le propre d'un cas d'usage du *Big Data* est de jeter des ponts entre différents types de data (digitales, externes, financières, ressources humaines, structurées et non structurées…). Or, dans vos systèmes informatiques existants, toutes ces data sont en silos, prisonnières des *datasets* (base de données) métiers, dans lesquels elles ont été initialement collectées. L'enjeu technologique majeur ? Rassembler toutes ces data en un seul et unique endroit. Nous appellerons cet endroit le *datalake* (lac de données, en français).

DES DATA CONSOMMABLES EN TEMPS RÉEL

Vous l'avez lu dans la partie «La data, pour mieux faire!» (p. 3-12), le *Big Data* est étroitement lié au digital, à l'activation chaude. Vos data doivent donc pouvoir être consommées, en temps réel, par vos interfaces Web, mobiles ou applicatives. C'est précisément (et uniquement) cette connexion avec vos outils opérationnels existants qui donnera toute leur valeur à vos algorithmes. Il vous faut donc un *datalake*, mais surtout un *datalake* connecté au reste de votre écosystème informatique en place.

À présent, dessinons un modèle générique (donc simpliste) d'architecture *Big Data*. Le Grand Soir informatique évoqué, c'est d'abord la mise en place de trois éléments techniques indispensables.

Copyright © 2018 Dunod.



- **D'abord, le *datalake*.**
  C'est votre poumon data. Il s'agit, ni plus ni moins, d'une base de collecte, mais qui s'appuie sur tout un lot de technologies plus récentes, indispensable pour accueillir trois nouvelles sources de data (digitales, comportementales et externes). De même pour la plupart des data non structurées, à très fort volume. Ce *datalake* vous permettra également d'y faire tourner des modèles de calculs sophistiqués, et rapidement. À noter que la plupart des technologies derrière ces infrastructures sont *Open Source*[1], ce qui, rappelons-le, n'est pas synonyme de gratuit. Citons le *stack*[2] Hadoop, le plus répandu et le plus iconique de l'écosystème.

1. Désigne des logiciels dont le code est mis à la disposition du grand public.
2. Désigne un ensemble cohérent de logiciels qui prennent sens dans l'écosystème d'un éditeur.

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

QUEL AFFREUX JARGON !

HADOOP

Désigne un vaste écosystème de technologies informatiques qui permet de distribuer un volume important de data sur plusieurs fichiers différents, en réduisant considérablement le temps de calcul. Sur la base d'une architecture partagée, un grand nombre d'outils et logiciels sont disponibles : HBase, Hive, Pig, NewSQL… pour rechercher, optimiser et requêter vos data !

- **Puis, les connecteurs.**
  Le *datalake* se nourrit de vos data. Plus elles sont récentes, plus vos modèles statistiques et vos outils opérationnels qui les consommeront seront fiables. Nous vous conseillons donc, lors de ce Grand Soir informatique, deux manœuvres. D'abord, ne vous contentez pas d'exports manuels de data, mais bâtissez, à défaut de véritables API[1] (voir p. 67), des exports automatiques, les plus fréquents possibles, depuis vos systèmes informatiques. Puis, veillez à ne pas récupérer, pour un système donné, les seules data nécessaires aux cas d'usage en cours… mais bien l'ensemble des data accessibles dans ce système. Elles pourront, demain, vous être bien utiles pour de prochains cas d'usage ! Ne raisonnez pas à court terme, sinon vous vous priverez d'économies d'échelle, concomitantes à ce type de plateformes.
- **Enfin, les fenêtres.**
  Nous l'avons répété, vos data n'ont de valeur qu'à condition d'être utilisées, donc d'être accessibles. Pour ce faire, votre *datalake* ne doit pas rester fermé sur l'extérieur. Ouvrez donc des fenêtres, pour deux usages principaux. D'abord celui de vos statisticiens, pour leurs calculs, segmentations et autres modèles. Il leur faudra

1. *Application Programming Interface* – des détails plus loin dans ce chapitre.

Copyright © 2018 Dunod.

des logiciels statistiques adaptés[1], directement branchés sur votre *datalake*. Puis, celui de vos outils opérationnels, qui doivent pouvoir consommer, en temps réel, la data, en lien direct avec votre *business*. Pour ces connexions, ce sont des API qui seront nécessaires.

EN VOILÀ UNE BONNE QUESTION ! Faut-il mettre mon *datalake* dans le *cloud* ?

C'est une tentation, loin d'être honteuse. Pour travestir la fameuse citation de J. M. Keynes : *À long terme, tous les systèmes seront sur le cloud*. Le *cloud* désigne le fait d'héberger vos data dans des serveurs localisés hors de votre entreprise, et administrés par des fournisseurs tiers (Amazon, Google, Microsoft, pour citer les plus fameux…). En ce qui concerne le *datalake* : suivez la stratégie globale de *cloud* de votre organisation. Pour le *Big Data*, rien ne justifie d'aller plus vite, ou de faire autrement. La plupart de vos outils digitaux (A/B *testing*, gestion de sites Web, outils d'enchères publicitaires…) fonctionnent déjà sur le *cloud*. Une réflexion, plus largement : si votre entreprise est soumise à d'intenses et de fréquentes cyberattaques, le *cloud* est un élément de réponse. Il est illusoire de penser que vos data, parce qu'elles sont hébergées dans vos murs, sont plus en sécurité qu'à l'extérieur. C'est plutôt le contraire qui est vrai. Les leaders du *cloud* ont justement la taille critique (que vous n'avez peut-être pas) pour développer les défenses technologiques les plus robustes contre les *hackers*.

Ces fondations technologiques bâties, et l'investissement de départ consenti, tous les cas d'usage data supplémentaires vous coûteront marginalement moins. Et vous bénéficierez (nous l'espérons) de rendements *business* croissants.

Dans votre brief initial aux équipes informatiques, précisez-leur un élément majeur. Où souhaitez-vous mettre le poids du corps, l'analytique ou le transactionnel ? Précisons ce que recouvrent ces deux notions.

1. Pour les plus initiés, de type R ou SAS.

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

L'ANALYTIQUE

Parions qu'une grande partie de vos cas d'usage data sont de nature plutôt analytique, c'est-à-dire que vous acceptez que les data, les segmentations ou les prédictions issues de vos algorithmes ne soient pas fiables à 100 %. Vous acceptez donc une dose de probabilité. Cela concerne tout particulièrement les cas d'usage marketing. Pour eux, pas indispensable d'avoir une data sûre et certaine, car les leviers d'activation, en aval, sont probabilistes : achat média, marketing direct…

Dans cette configuration, l'architecture *Big Data* présentée ci-dessus pourra être relativement déliée des systèmes informatiques. Elle tolère une moindre fiabilité des data. Les équipes informatiques pourront donc aller plus vite dans la mise en place des fondations technologiques.

LE TRANSACTIONNEL

A contrario, certains cas d'usage sont par nature transactionnels. Citons quelques exemples : un enrichissement d'adresses e-mails sur un fichier client ; un modèle de rémunération ou de facturation ; une détection d'anomalies dans une chaîne d'assemblage industrielle… Ces cas ont une double caractéristique. Les data en question sont critiques pour le bon fonctionnement des systèmes opérationnels. La fiabilité de ces dernières doit donc être à 100 % (sauf *bugs*). Dans cette configuration, l'architecture *Big Data* sera plus proche des systèmes informatiques existants. Vos équipes mettront plus de temps à la bâtir, mais le système sera plus robuste, en termes de récence et de fiabilité des data. Enfin, si vos cas d'usage data sont presque tous transactionnels, vous pourrez privilégier, en lieu et place du *datalake* mentionné, la mise en place d'un autre outil : le *Master Data Management* (MDM). Contrairement au *datalake*, ce dernier abrite des data plus structurées et une liste de fonctionnalités sur l'étagère, pour mieux s'interfacer avec vos autres outils opérationnels, déjà en place. Un MDM, par ailleurs, tient lieu d'excellent fichier client !

Copyright © 2018 Dunod.

Loin de nous l'idée d'opposer l'analytique au transactionnel. Il s'agit de placer le curseur au bon endroit entre ces deux dimensions, laquelle sera majeure, laquelle sera mineure pour vous. Dans les organisations dont les data, de base, sont dans un très piteux état, l'étape transactionnelle de remise à niveau s'imposera. Pour des organisations dans lesquelles la valeur *business* est d'abord tirée par les leviers marketing, c'est plutôt par l'étape analytique qu'il faudra débuter. À vous de juger !

> **La vérité n'est pas l'exactitude.**
>
> **Henri Matisse**

La bonne santé data de votre entreprise se mesurera, au-delà de la réussite de ce Grand Soir technologique, à la qualité de vos data dans le temps, et donc à votre capacité à faire respecter une vraie discipline opérationnelle dans vos équipes et vos projets. La clé : les fameuses API ! Ce sont des interfaces de programmation qui permettent à deux de vos machines d'échanger des informations, dans un langage commun et partagé. Elles garantissent l'interopérabilité de vos systèmes informatiques, en interne comme en externe. Vous n'insisterez jamais assez sur leur importance.

QUEL AFFREUX JARGON !

API (OU APPLICATION PROGRAMMING INTERFACE)

Sans aucun doute l'acronyme informatique le plus important à connaître.

Il désigne un ensemble normalisé de classes, de méthodes ou de fonctions qui sert de façade, par laquelle un logiciel offre des services à un autre logiciel.

Concrètement, une API est une fenêtre ouverte sur un système informatique, ainsi qu'un langage qui permet à ce système d'être compris par d'autres systèmes, pour échanger des informations. Si votre entreprise n'a pas encore investi dans le développement d'API, c'est un très mauvais présage pour le succès de votre transformation ! Pressez le pas !

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

Plaçons-nous dans les pas (intimidants) de Jeff Bezos, le patron d'Amazon. Pour ses collaborateurs, il est allé jusqu'à édicter trois principes non négociables sur les API.

D'abord, toute data, chez Amazon, doit être rendue accessible *via* une API.

Puis, tout échange de data entre les équipes doit passer exclusivement par ces API : finis les e-mails, les fichiers Excel et autres exports.

Enfin, toute API doit pouvoir être externalisée, c'est-à-dire être utilisée par des interfaces et des développeurs tiers, lorsque Amazon le souhaite.

À l'instar d'Amazon, tout projet informatique dans votre entreprise doit porter une part d'API. Au-delà des bénéfices tangibles pour la data, elles sont un prérequis à toute digitalisation sérieuse de votre activité. Sans API, vos multiples interfaces (applications mobiles, sites Web…), seront aussi rigides que les systèmes informatiques « en cuisine » sur lesquels elles s'appuient. Et vous perdrez, de fait, la vélocité indispensable. Enfin, les API sont une condition de l'ouverture de votre entreprise sur l'extérieur. Si Google n'avait pas ouvert les API de son outil de cartographie Google Maps, Uber n'aurait jamais pu lancer son service de transports. Demain, pas d'écosystème vertueux de partenaires, sans APIsation de leurs systèmes informatiques. Après-demain, une grande partie du PIB national s'appuiera sur les API et leur usage. En 2018, 50 % des revenus d'Amazon seront déjà liés à leur consommation ; 60 % dans le cas d'eBay, et 80 % dans le cas d'Expedia… et vous ?

Copyright © 2018 Dunod.

2 | VOIR GRAND, TRÈS GRAND !

EN VOILÀ UNE BONNE QUESTION ! Comment vérifier que nous n'avons pas jeté de l'argent, beaucoup trop d'argent, par les fenêtres ?

La data doit faire bouger les lignes. Au risque, dans le cas contraire, de rester anecdotique, et de ne pas justifier les investissements financiers et humains consentis. Facile à dire, plus difficile à mesurer. Pourtant, c'est la rançon de la gloire. Sur la longue route, votre trajectoire data ne sera crédible qu'à condition d'y accrocher des gains *business* sonnants et trébuchants. Néanmoins, pas question de passer des heures sur des calculs d'apothicaires, ou d'hypothétiques *business cases*. Voici deux méthodes pragmatiques (et plutôt simples) à activer, pour mesurer l'impact *business* de vos cas d'usage data. À noter : nous ne parlons pas des entreprises publicitaires, qui vivent directement de la monétisation de leurs data, et pour lesquelles la question de la valorisation de la data est directement celle de son résultat d'exploitation !

PROUVEZ, PAR A + B, L'APPORT BUSINESS DE LA DATA

C'est la configuration idéale, mais, malheureusement, elle n'est pas adaptée à tous les cas d'usage. Pour prouver, sans contestation, l'apport d'une data, vous aurez recours à des méthodologies de type A/B *testing*. Pour plus de détails, rendez-vous à la partie « La data, ce n'est pas une arme ! » (p. 33-40). Ces méthodes neutralisent beaucoup de biais et permettent de répondre à la question : *La même action, avec ou sans cette data, ça change quoi ?* Les cas d'usage mixant data et digital, notamment publicitaires, s'y prêtent tout particulièrement, car les outils d'activation associés intègrent nativement des fonctionnalités d'A/B *testing*.

Utiliser des data sociodémographiques pour cibler plus précisément mes publicités sur Google, ça paie ? Les e-mails envoyés à mes clients détectés comme fragiles ont-ils un impact sur leur fidélisation ?

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

Le recours à l'A/B *testing* fonctionne aussi pour certains cas d'usage touchant les plateformes téléphoniques. *La détection d'un moment de vie améliore-t-elle l'efficacité de mes appels sortants ?*

Voici la règle : s'il vous est possible, dans le « dernier mètre » de l'activation, de décider d'intégrer ou pas la data, il devient possible de démontrer son apport, sans contestation. Attention, n'abusez pas de l'A/B *testing*. Pour une bonne raison : cela prendra du temps à vos équipes.

ESTIMEZ L'APPORT BUSINESS DE LA DATA

Certains cas d'usage bouleversent tellement les procédures opérationnelles que la comparaison A/B *testing* n'a plus lieu d'être. Prenons un exemple : pour détecter une fraude client, vous mettez en place une toute nouvelle procédure d'alerte pour vos conseillers, branchée sur votre *data-lake*. Comment comparer la procédure avec ou sans data, puisqu'il n'y a qu'une seule procédure (celle que vous venez de mettre en place), et que la data en est justement partie intégrante ?

Faites le deuil d'une démonstration parfaite, et adoptez une méthode plus pragmatique, celle de la preuve par la contribution. Qu'est-ce à dire ? La data est d'abord un facilitateur, une capacité que vous activez, au service d'enjeux *business* plus larges, et qui la dépasse : le multi-équipement, la fidélisation, le développement de nouveaux services, la qualité du recrutement… Dans ce cas, la data n'est qu'un maillon de la chaîne, aux côtés de nombreux autres leviers activés. L'enjeu est d'estimer (avec discernement) sa contribution.

Prenez les cas d'usage qui n'existeraient pas sans le recours à la data, pour lesquels cette dernière est une condition *sine qua non* de fonctionnement. Pour ceux-là, la contribution attribuée à la data sera naturellement supérieure à 50 %. Il y a aussi les cas d'usage pour lesquels la data n'est qu'une force d'appoint, un effet *boost*. Voilà une question à poser au sponsor, pour affiner l'estimation : *Sans l'apport de la data, qu'auriez-vous fait différemment ?*

Copyright © 2018 Dunod.

Une précision importante : pour valoriser la data, ne vous arrêtez pas uniquement à la valeur d'un algorithme donné, ou à la présence d'une data spécifique. Ayez le souci de bien élargir la mesure à l'ensemble des procédures qui permettent de faire quelque chose de la data, de la lier à vos outils opérationnels ou à vos utilisateurs finaux. Parfois, c'est simplement le fait de rapprocher des data appartenant à deux systèmes informatiques particulièrement cloisonnés, *via* votre *datalake*, qui déclenchera l'étincelle. Dans d'autres cas, c'est une toute nouvelle connexion plus rapide entre un *datalake* et un outil opérationnel qui fera la différence. Sans forcément utiliser des data supplémentaires, la clé des gains *business* est souvent une affaire de tuyaux !

> **From this day to the ending of the world ; But we in it shall be remembered ; We few, we happy few, we band of brothers.**
>
> **Henri V**[1]

Nous avons clarifié (espérons-le) la méthode pour mesurer les gains *business* issus de la data. Il vous reste à déterminer votre ambition globale. Car vos équipes en ont besoin, d'autant plus sur un territoire nouveau, peu connu. Tel John F. Kennedy en 1961, qui promettait d'envoyer des Américains sur la lune avant 1970, il vous faudra fixer un cap, le plus haut possible. Pas simplement pour la beauté du geste, mais parce qu'un objectif de gains ambitieux oblige vos équipes à réfléchir différemment.

Bien sûr, la méthode *bottom-up* a également son utilité. Elle vise à interroger directement vos équipes sur ce qu'elles pensent générer comme valeur avec leurs initiatives data. Mais attention : au moment de l'addition finale des contributions, le résultat est décevant. N'oubliez pas que le changement n'est pas naturel. Et, comme le digital, la data, c'est de la transformation.

1. « À compter de ce jour jusqu'à la fin du monde, sans que de nous on se souvienne de nous, cette poignée, cette heureuse poignée d'hommes, cette bande de frères. »Shakespeare, *Henry V*, trad. Jean-Michel Déprats, Gallimard, coll. « Folio théâtre », 1999.

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

Citons le fameux mot d'Henri Ford : *Si j'avais demandé à mes clients ce qu'ils attendaient, ils auraient répondu un cheval plus rapide.* Assumez donc une part de *top-down* dans la fixation de l'ambition. Sortez du bois, en fixant une cible de gains à trois ans très élevée, *a minima* trois fois la somme des promesses de gains récoltés dans la démarche *bottom-up*. Vous trouverez facilement, dans un second temps, la bonne clé de répartition pour allouer cette ambition globale de gains, entre vos différentes équipes.

ILS L'ONT FAIT !

Google est l'archétype de l'entreprise qui voit grand.

Elle est allée jusqu'à faire de cette démesure une véritable discipline opérationnelle, au sein de ses projets internes qui touchent ses produits. À chaque nouvelle initiative candidate à un financement conséquent, le *top management* de la firme californienne lui fait passer le test dit du « brossage de dents » : *Est-ce que ton produit va toucher tous nos utilisateurs, au moins deux fois par jour ?… Stress test* exigeant, s'il en est !

Copyright © 2018 Dunod.

ILS NE L'ONT PAS FAIT...

Des sujets à fort potentiel, l'État n'en manque pas ! Citons la fraude aux prestations sociales : elle pèserait entre 2 et 3 milliards d'euros par an en France. Ou encore la fraude à l'Assurance Maladie. Chaque année, 163 milliards d'euros de prestations sont versés aux Français. Sur ce total, seulement 245 millions d'euros ont été détectés comme frauduleuses en 2016, soit une progression quasi-nulle par rapport à 2015. C'est un taux de détection très bas. Alors que les data sont nombreuses. Une « simple » étude de la rémunération des praticiens permettrait, par exemple, en la segmentant par rapport au lieu, à l'activité et au profil de leur patientèle, d'identifier un bon nombre d'anomalies.

Rêvons un peu : et si le ministre de la Santé fixait comme objectif à ses équipes : *Trouvez, d'ici 2020, 3 milliards d'euros de fraude !*

Espérons alors que quelques premiers *Data Scientists* seraient recrutés au ministère…

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

3 | HUMAIN, TROP HUMAIN !

EN VOILÀ UNE BONNE QUESTION ! Comment prioriser entre toutes nos initiatives data ? Par quoi démarrer ?

Vos objectifs *business* sont ambitieux. Les *Data Scientists* que vous aurez envoyés au contact des autres directions ont stimulé leur appétit data. En réaction, vos équipes ont rivalisé d'imagination. Les cas d'usage data les plus divers sont proposés : des ressources humaines à la logistique, en passant par les fonctions marketing. Tout votre petit monde y va de ses bonnes idées. Vous voici, à présent, devant un autre challenge : la nécessité de trancher et de prioriser, en prenant l'hypothèse (probable) que vous êtes soumis à un jeu de contraintes financières. La clé de priorisation par les gains *business* attendus paraît, à première vue, la plus naturelle. Nous avons couvert le sujet dans la précédente partie « Voir grand, très grand ! » (p. 69-73). Elle ne suffit pas. Filtrez les cas d'usage proposés à l'aune d'un autre critère : la capacité des équipes à délivrer ces gains et leur probabilité de succès dans l'exécution. Voici une question structurante pour bien l'apprécier : *Quelle part d'humain dans le cas d'usage data ?*

LES CAS D'USAGE « HUMAIN »

L'humain fera le succès de l'opération. C'est le cas lorsqu'au bout de la chaîne, un utilisateur final doit s'emparer de la data, l'interpréter et effectuer une action, indispensable pour récolter les gains *business* attendus.

L'exemple le plus parlant est celui de la distribution. Vous pourriez, *via* un algorithme data sophistiqué, parvenir à anticiper l'infidélité de certains de vos clients. Mais c'est, *in fine*, votre conseiller de clientèle qui, après avoir réceptionné votre alerte dans son poste de travail, devra décrocher son téléphone, parler avec le client et trouver les mots pour le convaincre de rester chez vous. Dans ce cas-là, la data sert de déclencheur, mais c'est l'humain qui fait le succès ou l'échec de l'activation, en bout de chaîne.

Copyright © 2018 Dunod.

Prenons un autre secteur, celui de la santé : le célèbre moteur d'intelligence artificielle Watson, construit par IBM, apporte des éléments précieux de contexte pour aider un médecin dans son diagnostic. Mais le dernier mot, le diagnostic final, c'est le médecin qui l'exprime, et qui en porte la responsabilité.

Si les progrès rapides de l'automatisation vont probablement modifier la place de l'humain dans le futur, l'émotion, l'expertise et l'intuition garderont un rôle important. Dans vos cas d'usage data également. Qu'est-ce à dire ? Que lorsqu'une part importante d'humain est impliquée, les gains *business* à la clé sont souvent plus élevés, mais plus difficiles à aller chercher. Qui dit humain dit friction, besoin de formation, et une courbe d'apprentissage plus longue.

Un conseil : pour démarrer votre trajectoire data et obtenir des succès rapides et démontrables, ne commencez pas par ces cas d'usage « humain ».

C'est choisir la face nord.

LES CAS D'USAGE « MACHINE »

La machine fait l'essentiel, c'est le cas inverse de celui qui précède. Tout au long de la chaîne opérationnelle, la part de l'humain est ici réduite à sa plus simple expression. Quelques illustrations : l'optimisation d'une chaîne d'approvisionnement de stocks ; le *pricing* dynamique de billets d'avion ; l'achat publicitaire digital sur Google ; le calcul des provisions techniques dans un bilan comptable…

Dans cette configuration, la partie est plus facile. La machine n'ayant pas d'états d'âme, elle se contente d'exécuter une série d'opérations successives. Dans votre entreprise, un niveau moyen élevé d'automatisation des procédures est d'ailleurs un signal de fort potentiel *business* pour la data, comme vous l'avez lu dans la partie « La data, pour mieux faire ! » (p. 3-12). Moins de risque de frictions humaines, mais davantage de risque de *bugs*. Privilégiez ces cas d'usage « machine » au

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

démarrage de votre voyage, car les gains sont plus rapides, plus probables, et surtout plus facilement démontrables.

Autre avantage des cas d'usage « machine » : la boucle de rétroaction est plus rapide. En effet, un algorithme data se nourrit de la réalité. Il a besoin d'avoir du *feed-back* sur la qualité de sa prédiction. *Était-elle bonne ? Sinon, quelle était la bonne réponse ?*

C'est que nous appelons le besoin de labellisation. Plus que l'humain, la machine peut répondre, dans le temps chaud, sur le succès de l'activation data, *via* un taux de clic sur une bannière publicitaire, un écart à la prédiction de stocks ou d'appels sur une plateforme téléphonique… Plus votre algorithme obtiendra de retours du réel, plus il progressera vite !

QUEL AFFREUX JARGON !

LABELLISATION

Processus indispensable à son apprentissage continu qui consiste à informer un algorithme du succès de sa prédiction, c'est-à-dire de l'écart entre cette dernière et la réalité. Prenons l'exemple de la détection d'une fraude client : l'algorithme a besoin, pour prédire, d'avoir une série de data labellisées, c'est-à-dire une suite de cas où la fraude client est avérée. Sans labellisation, la performance d'un algorithme est indémontrable. Le plus puissant, c'est lorsque la labellisation se fait dans le temps chaud de l'activation, comme dans les cas d'usage « machine ». L'algorithme peut alors optimiser, en continu, la précision de sa prédiction.

Copyright © 2018 Dunod.

ILS L'ONT FAIT !

Le succès foudroyant des GAFA (Google, Amazon, Facebook, Apple), en termes de data, tient, pour une large part, à la combinaison de deux forces : l'extrême niveau d'automatisation de leurs procédures, ainsi que la haute fréquence du *feed-back* utilisateur et donc le niveau très élevé de labellisation associé.

Prenez le chiffre d'affaires de Google, aux alentours de 66 milliards de dollars, et comparez-le au nombre de ses collaborateurs, à peine 50 000 en 2017. Ou Facebook, dont le ratio est comparable : 17 milliards de dollars de chiffre d'affaires, pour 20 000 collaborateurs.

Ces jeunes entreprises sont davantage consommatrices de serveurs informatiques que de ressources humaines. Leurs cas d'usage data sont quasi uniquement « machine ». Toute modification de l'algorithme de Google a un impact immédiat sur la performance de son moteur de recherche. Pas de friction entre le code des ingénieurs en amont, et les utilisateurs derrière leurs écrans, en aval.

La puissance de ces acteurs tient aussi à la chaleur du *feed-back* utilisateurs. Pour Google, chaque clic sur son moteur de recherche est une sanction positive de la pertinence du résultat affiché. Pour Facebook, ce serait plutôt le *like* et sa nouvelle palette d'emojis, pour signifier la pertinence d'un *push* sur le fil d'actualisation. L'algorithme peut donc apprendre en temps réel.

En termes d'intelligence artificielle également, les deux acteurs californiens tiennent la corde. Les milliards d'utilisateurs de leurs services éduquent chaque seconde leurs robots. Sur Facebook, en taggant vos photos de vacances, vous lui apprenez à deviner, de lui-même, le contenu des images. En cliquant sur des résultats Google, vous lui donnez de précieuses indications sur la compréhension de la sémantique des phrases tapées en langage naturel sur son moteur de recherche. Grâce à vous, aujourd'hui, Facebook est déjà capable (tout seul) de distinguer un labrador d'un doberman sur une photo.

Et Google, de faire la différence entre les deux questions « *quel temps à Ibiza* » et « *quel temps pour Ibiza* »…

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

Dans l'exercice de priorisation des cas d'usage « humain » comme « machine », exigez que le sponsor (des détails dans la partie « Bousculer sans braquer ! », p. 81) formalise, et avec précision, les leviers d'activation associés. Détecter le moment de vie « déménagement » de vos clients, susceptible de déclencher une action de multi-équipement, c'est très bien. Mais ce n'est qu'une partie du chemin. Que faire de ce moment de vie, ensuite, pour toucher le client final ? Est-ce qu'il sera transmis à un conseiller en plateforme téléphonique, ou bien déclenchera-t-il l'envoi d'un e-mail ?

En fonction des leviers d'activation sélectionnés, le gain *business* issu de la data variera du tout au tout. Derrière chaque cas d'usage, ayez le souci d'estimer l'ensemble des coûts associés, et pas uniquement les coûts liés au traitement de la data. D'expérience, c'est précisément le déploiement de ces leviers d'activation classiques (forces commerciales, investissements marketing…) qui coûtent le plus cher… et de loin ! Bonne nouvelle : sur ces territoires, vous connaissez les prix.

Copyright © 2018 Dunod.

4 CHAPITRE Libérer vos énergies

Vous avez une stratégie data ambitieuse. Vous savez dans quelle direction faire bouger votre culture d'entreprise. À présent, c'est une nouvelle équipe «noyau dur» qu'il vous faut constituer. Pourquoi ? Parce que le *Big Data*, en tant que tel, est une discipline nouvelle, qu'il vaut mieux concentrer dans votre organisation, pour éviter de la diluer trop vite. Vient d'ailleurs le moment de lui donner une définition, au *Big Data*. Nous le caractériserons par l'explosion de trois nouvelles sources de data, non structurées (la définition dans la partie «La data, pour mieux faire!», p. 3-12) : les data digitales (liées à notre consommation des écrans), les data comportementales (liées à notre usage des objets connectés, à notre géolocalisation…) et les dates externes (liées à l'*Open Data*, aux échanges entre entreprises…). Deux enjeux : un enjeu informatique et un enjeu statistique. Un enjeu informatique, car vos systèmes d'information classiques (les fameux *datasets*) ne peuvent pas accueillir ces nouvelles data (pour l'explication, rendez-vous à la partie «Le Grand Soir technologique, c'est ce soir!», p. 61). Puis un enjeu statistique : les modèles traditionnels, basés (nous caricaturons) sur un faible volume de data intelligibles, et sur de la causalité, laissent place à des modèles nouveaux, beaucoup plus consommateurs en data, et qui s'accommodent très bien de la corrélation. Autrement dit, votre équipe «noyau dur» sera composée d'ingénieurs et de statisticiens, mais d'un genre différent de ceux dont vous avez l'habitude. Ce chapitre est là pour vous aider à la bâtir, et lui donner toutes les chances pour réussir.

Copyright © 2018 Dunod.

Copyright © 2018 Dunod.

1 | BOUSCULER, SANS BRAQUER !

EN VOILÀ UNE BONNE QUESTION ! Comment bousculer notre organisation, sans la braquer ?

Comme nous l'avons déjà dit, la data est là pour faire bouger les lignes. Au commencement de votre aventure, marquez une accélération, et une accélération visible. Pour ce faire, il vous faudra très probablement recruter des talents en externe. L'étude IDC déjà citée (p. 3) montre une prise de conscience : 80 % des Directeurs de Systèmes d'Information et 90 % des Directeurs Marketing pensent qu'ils ne peuvent pas mener de projets *Big Data* avec leurs seules compétences internes.

Ces nouvelles expertises, logez-les dans une équipe transverse, rattachée à la Direction Générale. Avec, comme boussoles, des objectifs data spécifiques, traduisant l'ambition que vous aurez donnée au sujet.

Plusieurs vertus à ce positionnement transverse. D'abord, éviter la rivalité classique de territoires entre les équipes marketing, *business* et informatique. Si possible, épargnez-vous ces frictions classiques, surtout pour démarrer.

Autre avantage majeur : éviter à cette nouvelle équipe *Big Data* d'être écrasée, dès le début, par la pression des objectifs de rentabilité à court terme. Loin de nous l'idée que le *Big Data* doive échapper à l'exigence d'un retour sur investissement, problématique que nous avons d'ailleurs abordé en détail dans la précédente partie « Humain, trop humain ! » (p. 74-78). Non, le *Big Data* doit générer des gains *business*. Mais il est illusoire d'en exiger un retour sur investissement dès les premiers mois. Il y a une phase d'investissement informatique minimale, qui ne pourra être amortie que sur la moyenne route, au fil de vos cas d'usage.

Raisonnez ici par analogie avec des transformations passées : l'informatisation, l'innovation ou encore le digital. Pour réussir ces tournants, mieux vaut isoler l'équipe en charge, au démarrage, des autres équipes. Et, à mesure du progrès collectif, rattacher, puis dissoudre la nouvelle

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

expertise dans les directions en place. Voici, caricaturées, les trois étapes successives (et selon nous incontournables) d'une transformation.

D'ABORD, LE RÉVEIL COLLECTIF

C'est le début. Votre organisation perçoit une rupture technologique majeure sur le marché, et comprend qu'elle doit changer et anticiper. Pendant plusieurs mois, elle fait feu de tout bois, sans considération spécifique pour le retour sur investissement des initiatives lancées.

Votre *top management* est invité à faire un petit tour sur la côte ouest américaine pour s'encanailler avec GAFA et *start-up* (un voyage pudiquement baptisé *learning expedition*). Des outils flambant neufs sont bâtis, à grands renforts d'investissements : un *datalake*, des applications mobiles, des prototypes de nouveaux produits…

Vous expérimentez, dans l'allégresse générale, et enchaînez *Proofs of Concept* (POC) sur POC. De vastes programmes de formation sont lancés à destination des collaborateurs, pour les rassurer (à juste titre) sur l'avenir.

En termes d'organisation, vous réalisez vos premiers recrutements externes, de façon à faire venir de l'expertise différente. Une équipe « noyau dur », dédiée à la transformation, est constituée, généralement rattachée à la Direction Générale, pour qui c'est l'occasion de se mouiller sur le sujet.

Quelques articles sont publiés, dans la presse spécialisée ou généraliste, pour dire que l'entreprise bouge.

Et, à vrai dire, dans cette première étape du voyage, indispensable, c'était bien l'objectif principal : bouger !

ENSUITE, L'ADOPTION COLLECTIVE

Quelques mois après votre démarrage en trombe et les premières vagues d'investissements passées, vient le temps de l'adoption. Votre équipe « noyau dur » a parfaitement joué son rôle de poil à gratter, en bousculant, tant que faire se peut, les équipes en place.

Copyright © 2018 Dunod.

L'exigence du retour sur investissement commence à poindre le bout de son nez. Progressivement (et plus ou moins subtilement), il est demandé à l'équipe de normaliser son fonctionnement, de le formaliser, et de se donner des premiers objectifs *business*. *Le digital, c'est bien beau, mais comment cela impacte-t-il le bilan ? Le Big Data, ça fait combien de revenus en plus ?*

C'est le temps aride des premiers *business cases*, pour accrocher la transformation au reste de l'activité. L'équipe « noyau dur » reste souvent unifiée, mais elle est souvent réorganisée de l'intérieur, pour mieux coller à ses nouveaux objectifs : recherche et développement, acquisition de prospect, fidélisation du client…

ENFIN, LA DISSOLUTION COLLECTIVE

Au bout de plusieurs années, l'innovation, le digital, la data, ont terminé de se normaliser. Parce que la greffe a pris, l'activité de votre entreprise a été impactée positivement et s'est transformée.

Il est temps, pour l'équipe « noyau dur » des premières batailles, de rejoindre définitivement le vaisseau amiral. Dans la période précédente, elle s'était organisée autour d'objectifs *business*. C'est donc très naturellement que chaque élément trouve sa place dans les directions existantes.

La distribution digitale devient un bout de la distribution classique ; une partie de l'équipe *Big Data* rejoint les équipes CRM et de *business intelligence* ; ceux qui avaient lancé le tout dernier produit innovant continuent le travail dans la direction (plus classique !) du marketing produit…

Sur la question *Big Data* (les *start-up* et les *pure players*[1] mis à part), peu d'organisations ont dépassé la première étape, celle du réveil collectif. Sur le digital, la plupart (heureusement) sont parvenues au

1. Désigne les acteurs économiques exerçant leur activité uniquement sur Internet.

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

second stade. Le chemin de maturité est long. Si cette découpe est caricaturale, elle possède sa part de vérité. La data n'est pas votre cœur de métier ? Mieux vaut concentrer et séparer la nouvelle équipe du reste de l'entreprise, et la rattacher à la Direction Générale. La rattacher à une équipe *business* déjà en place, c'est véhiculer l'idée que la data est l'affaire d'une équipe, alors qu'elle est l'affaire de tous. La rattacher à l'informatique, c'est laisser penser que le *Big Data* n'est qu'un outil informatique comme les autres, à disposition sur l'étagère, alors qu'il est, avant tout, une transformation de culture, de procédures et de modèles économiques (pour vous en convaincre, allez à la partie « Un projet informatico-informatique, tu éviteras ! », p. 109). Le fait de rompre les logiques classiques d'organigramme est un très bon signal. Parce que la matière est différente, il faut faire différemment.

QUEL AFFREUX JARGON !

CDO

Comme évoqué, l'organisation data est très similaire à celle du digital. Jusque dans la dénomination CDO. La lettre « D » marche en effet pour les deux périmètres : *Chief Data Officer* et *Chief Digital Officer*. Qu'est-ce que l'acronyme désigne, au-delà du terme à la mode ?

Commençons par la version négative… Acronyme à la mode, attribué généralement à un jeune trentenaire, recruté de l'extérieur, pour être l'étendard, interne comme externe, de la transformation d'une entreprise. Au quotidien, le bonhomme se retrouve souvent sans équipe ni budget… ce qui explique sa faible durée de vie (un ou deux ans)**… Mais préférons-lui la version positive !** Le CDO a une mission à la hauteur de ses responsabilités : convaincre, puis embarquer toute votre entreprise dans l'aventure data et digitale. Rattaché à la Direction Générale, disposant d'un budget et d'une équipe propre, il est au front des premiers cas d'usage data, pour démontrer leur valeur *business*.

Copyright © 2018 Dunod.

2 | LA DIVERSITÉ, AVANT TOUT !

EN VOILÀ UNE BONNE QUESTION ! Combien de *Data Scientists* me faut-il recruter ?

Cassons tout de suite une idée toute faite, et déjà très (trop !) répandue sur le marché : qu'une réussite *Big Data* se mesurerait au nombre de *Data Scientists* que vous réunissez autour de la table, pour « craquer » une problématique *business*. Pour preuve : prêtez attention aux forces en présence, parmi les leaders du marché. Chez les Google, Facebook et autres Palantir, les *Data Scientists* sont là, mais en nombre infinitésimal par rapport à la population des ingénieurs. Pourquoi ? Pour tirer des gains *business* de la data, l'essentiel se joue dans leur accessibilité, et dans l'efficacité et la fluidité de la chaîne opérationnelle en aval, pour l'activer.

La sophistication de l'algorithme compte peu (rendez-vous à la partie « Les algorithmes, tu ne sophistiqueras pas ! », p. 106). Certes, quelques points de performance grappillés par un *Data Scientist* sur la prédiction d'un moment de vie client sont appréciables, mais ils ne changeront pas les lignes. Tout particulièrement pour les cas d'usage « humain » (plus de détails dans la partie « Humain, trop humain ! », p. 74), votre succès tiendra à autre chose.

L'équipe *Big Data* « noyau dur » que vous constituerez sera donc un *patchwork*. Un alliage de compétences plus étendues, hors de la seule statistique. Nous parlons d'ingénierie, de *design*, d'architecture informatique… et la clé tiendra justement au bon dosage entre ces différentes disciplines. Le travail de la data, dans un cas d'usage « machine » comme « humain », se rapproche d'une chaîne d'assemblage industrielle. Parcourez-en les différentes étapes.

DES PLOMBIERS, POUR FAIRE CIRCULER LES DATA

Pour valoriser la data, encore faut-il y accéder. Or, les data remontent rarement à la surface de nos outils quotidiens. La plupart reste stockée

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

« en cuisine », dans vos systèmes informatiques, invisibles et sans grande vitalité. Poursuivez un rêve : *via* des API omniprésentes, que chacun de vos collaborateurs soit en capacité d'appeler n'importe quelle data, n'importe quand, sur n'importe quelle interface, en quelques clics. La réalité avec laquelle vous devez encore (malheureusement) sûrement composer est celle des systèmes informatiques archaïques, fermés sur l'extérieur, dont le fonctionnement, aléatoire, va jusqu'à ralentir la nuit.

Pour contourner cette rigidité, il vous faut une force de frappe d'ingénieurs. Des plombiers, au sens où ils vont manier de la tuyauterie pour fluidifier la circulation des data entre vos systèmes, vos outils et vos collaborateurs.

D'expérience, plus de la moitié du temps passé par vos équipes sur vos cas d'usage data tiendra à ce travail de plomberie. Pour ouvrir les systèmes informatiques, y intégrer des tuyaux, et programmer des exports de data plus ou moins fréquents.

Les ingénieurs *Big Data* doivent également posséder une fibre d'architecte, car il faut adapter la tuyauterie à la maturité (donc souvent la récence) de chaque système informatique (problématique traitée dans la partie « Le Grand Soir technologique, c'est ce soir ! », p. 61-68). Un peu comme l'alpiniste au pied du sommet qu'il doit gravir, il leur faut un plan de marche pour attaquer le système : *Quels formats d'exports ? Comment ne pas fatiguer la machine ? Quelle fréquence de communication des data ?*

Cette force de frappe d'ingénierie est surtout nécessaire au démarrage, pour traiter le stock. Sur le flux, moins de sueur : construisez des systèmes informatiques qui intègrent nativement les fameuses API !

DES SCRIBES, POUR NETTOYER LES DATA

Une fois le travail des ingénieurs derrière vous, imaginez les data présentes et disponibles. Mais ne vous précipitez pas ! Reste une étape indispensable, avant de les transporter dans le *datalake*, ou de les

Copyright © 2018 Dunod.

rendre accessibles à vos statisticiens : le nettoyage. Depuis des années, des décennies, les organisations fonctionnent essentiellement sur une tradition orale, quand il s'agit du traitement des data. *Cette information ? Je crois qu'elle est dans tel système, à côté de tel autre. À moins que non… attends, vas voir Sylvain, il te dira !*

Les référentiels, formels, décrivant la structure relationnelle entre les data, listant les méta-informations présentes sont, dans le meilleur des cas, incomplets… et dans le pire (mais le plus probable), tout simplement absents. Les choses sont dites, très peu écrites. Pour les malheureux collaborateurs qui s'intéressent à la data sous un angle différent, c'est un véritable travail d'archéologie qui est nécessaire !

C'est là que vous faites intervenir un nouvel acteur clé dans votre casting idéal, le *Data Manager*. Son rôle ? Formaliser des années et des années d'oralité. Procédant par cas d'usage data, il investigue, aux côtés des experts de chaque gisement de data, pour reconstituer des tableaux de data propres et intelligibles pour les potentiels utilisateurs.

Ce travail rend possible l'analyse du *Data Scientist* en aval. Mais pas seulement. Il permet également la mise en conformité réglementaire et juridique de vos différents cas d'usage. En effet, hors de question de se lancer dans une expérimentation data sans préciser ses finalités. C'est la position, sans ambiguïté, exprimée par la CNIL. Le *Data Manager* est donc le meilleur partenaire de votre équipe juridique.

Autre vertu de cette personne : elle est gardienne de votre *datalake*, à son lancement et dans le temps. Pas question de laisser passer de la « mauvaise data », dont la qualité est dégradée. Sinon, votre *datalake* se transformera rapidement en *swamp* (marais), jusqu'à ressembler à vos autres systèmes informatiques.

En bref : il vous faut, dans vos équipes, quelqu'un capable de répondre, à tous moments, à ces questions : *De quelle data s'agit-il ? D'où vient-elle ? Ai-je le droit de l'utiliser ?…* Et cette personne, c'est le *Data Manager* !

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

DES ALCHIMISTES, POUR METTRE DE L'INTELLIGENCE DANS LES DATA

Vient (enfin !) le temps de l'analyse, celui du *Data Scientist*. Il reste à ce dernier un peu de nettoyage. Pas celui que nous venons d'évoquer, celui qui porte sur la qualité intrinsèque des data.

Non, nous parlons, ici, de la transformation nécessaire pour les intégrer proprement dans un modèle statistique. Les data deviennent des *features*, à savoir des ensembles de variables regroupées entre elles, pour des raisons de simplification, et parce que considérées comme particulièrement utiles à la prédiction. Puis, après plusieurs simulations et ajustements, il s'agit pour le *Data Scientist* de coder le modèle le plus idoine. C'est une précision qui a son importance.

Le temps où, dans le travail de la data, il était possible de séparer l'informatique de la mathématique est révolu. Qu'est-ce qui explique cette évolution ? Beaucoup de glose a été produite sur le sujet. Disons que le rôle du statisticien a changé. Les premières entreprises qui embauchaient des statisticiens le faisaient avant tout dans le but de valider des hypothèses simples sur leurs produits. L'inventeur de la loi de Student travaillait dans une brasserie. Les modèles statistiques étaient peu complexes, tournés vers la description, plus que la prédiction (pour comprendre la différence, rendez-vous à la partie « Décrire et prédire, tu ne confondras pas ! », p. 123).

Ces dernières années, l'exigence vis-à-vis du statisticien s'est élevée d'un cran (en même temps que son salaire) : il doit prédire, donc ajuster constamment son modèle, et se porter garant des résultats dans la « vraie vie ». Un bon *Data Scientist* doit pouvoir produire son code, pour bien sécuriser l'implémentation de son modèle statistique. Cette exigence de maîtrise de bout en bout accentue le sens des responsabilités : un modèle statistique est conçu pour impacter l'activité, pas pour faire l'objet d'une publication de recherche fondamentale.

Rassurez-vous, les logiciels de type R proposent des librairies de modèles statistiques sur l'étagère. Ils économisent une bonne partie du travail de modélisation de vos *Data Scientists*.

Copyright © 2018 Dunod.

Même si un niveau minimum de customisation restera nécessaire de leur part !

DES PEINTRES, POUR REPRÉSENTER LES DATA

Une data n'a de sens que si elle est utilisée ! Placez-vous, ici, dans la configuration des cas d'usage « humain ». Pour réussir à être adoptée par des utilisateurs finaux, rendez la data simple et désirable.

C'est à dessein qu'est utilisé le terme de désirabilité. Nous ne sommes pas que des purs intellects, et nos émotions comptent beaucoup dans nos décisions, nos actions et nos réactions. Voilà donc une question systématique à vous poser : *Que doit ressentir l'utilisateur final au contact de la data ?* C'est là que le *designer* intervient, dans le dernier mètre (dont l'importance est creusée dans la partie « Le dernier mètre, tu réussiras ! », p. 99), pour, à l'écoute des utilisateurs finaux, réaliser des premières ébauches d'interface. Il s'agit d'y mettre des couleurs, de l'interactivité, voire une dose de jeu. Autant d'ingrédients, bien connus, de désirabilité. Vous connaissez l'importance de la *DataViz*, comme discipline (expliquée dans la partie « La data, ce n'est pas une arme ! », p. 33). Votre équipe *Big Data* doit accorder une place de choix à ces compétences de représentation. Ne les négligez pas : ils peuvent faire le succès ou l'échec de l'adoption !

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

Vous l'avez compris : la réussite de votre équipe *Big Data* tient au bon dosage de plusieurs compétences. Il est aussi probable que vous puissiez en dénicher parmi vos équipes internes. Et c'est positif. Il est dangereux de composer uniquement une nouvelle équipe sur la base de recrutements externes. La connaissance du métier est un atout majeur pour réussir. En la matière aussi, équilibrez ! Les parcours internes seront particulièrement utiles à vos ingénieurs et vos *Data Managers*, qui peuvent valoriser leur connaissance des systèmes informatiques et des data existantes. De la formation et de l'envie feront de certains de vos statisticiens déjà en place de très bons *Data Scientists* : laissez-leur leur chance !

Pour autant, la *Data Science* implique la maîtrise de nouveaux langages, de nouveaux logiciels et d'être à l'aise avec les data digitales. Veillez donc à recruter quelques talents en externe. Quant aux *designers*, vous en trouverez sûrement au sein de vos équipes digitales. C'est plutôt d'ailleurs une discipline de *free-lancers*, qui se prête à la prestation externe.

Une fois recrutés ces talents, fidélisez-les. Ne négligez pas l'importance de l'environnement de travail. Il y a de grandes chances que les nouveaux venus appartiennent à la génération Y. Quelques secondes de caricature : plutôt jeunes, ils sont davantage fidèles au projet qu'à l'entreprise (donc pas fidèles du tout !), en quête de sens, davantage attirés par des défis ponctuels que des postes régaliens. Préparez-vous à en perdre certains, après quelques mois. C'est la loi du genre, du marché du travail en l'occurrence.

Sur la data, les talents sont rares et les besoins des entreprises exponentiels. Variez donc un maximum les cas d'usage, laissez-leur expérimenter, acceptez le vagabondage quotidien sur des sites spécialisés, loin de vos problématiques *business* immédiates, mais nourriture indispensable à leur formation continue.

Un facteur clé de réussite du travail collectif, c'est le passage à l'agilité. D'accord, l'injonction à l'agilité tient de la vieille antienne. Elle char-

Copyright © 2018 Dunod.

rie beaucoup de promesses, beaucoup de fantasmes… et finalement de la déception. Car l'agilité comme discipline est finalement peu comprise.

Par agilité, nous désignons un ensemble de règles contraignantes, fixant le travail collectif : réunions quotidiennes, recours fréquent à la démo, modes de priorisation, regroupement physique sur un même plateau…

Point important : veillez à intégrer à votre dispositif des compétences précieuses, même si elles ne sont pas dans votre premier cercle data. Pensez aux équipes juridiques, ou encore à la sécurité informatique.

COMMENT ÇA FONCTIONNE, CONCRÈTEMENT ?

Quel est le bon dosage, justement, entre ces différents profils ? D'autant qu'en termes d'effectifs, rien n'est gratuit ! Voici une préconisation (à prendre avec beaucoup de pincettes) de répartition entre les profils, en fonction de quelques critères… si votre entreprise n'est pas déjà nativement digitalisée.

Le démarrage : il vous faut d'abord bâtir les fondations technologiques, et commencer à défricher vos data. Prenez une poignée d'ingénieurs (60 %), une cuillérée de *data managers* (30 %) et une pincée de *data scientists* (10 %).

Une majorité de cas d'usage « machine » : l'enjeu principal ? L'intégration informatique et la précision algorithmique. N'hésitez pas : gardez votre poignée d'ingénieurs (60 %), et rééquilibrez les doses entre les *Data Managers* (20 %) et les *Data Scientists* (20 %).

Une majorité de cas d'usage « humain » : il s'agit, ici, de faire jouer à plein la complémentarité data et collaborateurs. Aux côté de vos ingénieurs (50 %), ajoutez une même dose de *Data Managers* (20 %) et de *designers* (20 %). Puis pimentez avec un peu de *Data Scientists* (10 %).

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

Arrêtons-nous, pour terminer, sur la figure iconique du *Data Scientist*. Si elle n'est pas l'alpha et l'oméga de votre dispositif, elle en dit beaucoup du subtil équilibre à construire entre un état d'esprit « différent » et une compréhension *business* nécessaire. Un *Data Scientist* uniquement préoccupé de statistiques et de publications académiques bougera peu les lignes chez vous. *A contrario*, un *Data Scientist* uniquement soucieux de s'adapter à votre métier comme une éponge risque de perdre très vite sa fraîcheur avec des solutions à court terme. En résumé : il faut qu'il ait la curiosité du consultant pour le métier, sans son classicisme (voire sa rigidité) de pensée. Beau challenge !

> **« *Data Scientist*: The Sexiest Job of the 21st century. (Data Scientist : le job le plus sexy du 21e siècle.)**
>
> **D. J. Patil**

Copyright © 2018 Dunod.

3 | UN MANDAT CLAIR, SINON RIEN !

EN VOILÀ UNE BONNE QUESTION ! Comment donner de l'autorité à notre équipe *Big Data* en interne ?

Ce que vous cherchez à éviter : une équipe *Big Data* reçue avec bienveillance par les autres équipes de l'entreprise, qui passerait son temps dans des ateliers de travail fréquents et sympathiques, mais qui, sur le fond, ne serait pas crédible, ballottée et sans impact sur votre *business*. Sachez-le : le risque est élevé d'une gadgétisation de l'aventure *Big Data*. Pour l'éviter, soignez le mandat que vous donnerez à cette équipe, en précisant ce pour quoi elle a été constituée, et (encore plus important !) ce qu'elle a le droit de faire pour parvenir à l'objectif. La stratégie d'entreprise que vous avez conçue, la trajectoire data que vous avez déclinée, et les objectifs (ambitieux, on l'espère !) de gains *business* que vous avez fixés, devraient considérablement les aider. Voici quatre droits fondamentaux qui leur seront très utiles pour mener à bien leur mission :

- **Le droit de voir.**
  L'équipe *Big Data* doit pouvoir accéder aux data qu'elle souhaite, partout dans l'entreprise. À livre ouvert !
- **Le droit de comprendre.**
  L'équipe doit pouvoir être accompagnée par les métiers et les experts, pour comprendre les data : leur nature, leur finalité, les pincettes à prendre dans l'interprétation…
- **Le droit de challenger.**
  Elle doit pouvoir questionner, remettre en cause, et suggérer des façons alternatives de faire les choses, ou de répondre aux questionnements formulés par le sponsor. Il y aura beaucoup de naïveté dans leurs suggestions, mais aussi des choses à prendre.
- **Le droit de rencontrer.**
  Elle doit pouvoir rencontrer l'ensemble des acteurs ayant un rôle opérationnel impacté par le travail de la data (distributeurs,

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

gestionnaires, partenaires externes…). Notamment pour recevoir leur *feed-back*.

Le rattachement de l'équipe *Big Data* à la Direction Générale l'aidera à gagner de l'autorité en interne, mais ne suffira pas.

Arrêtons-nous quelques instants sur un rôle clé, celui du sponsor. Dans la trajectoire data que vous aurez conçue, chaque cas d'usage data porte une promesse de gains *business*. À cette promesse doit également correspondre une responsabilité métier (directeur de la distribution, directeur d'usine, directeur des ressources humaines…), claire et explicite. C'est le sponsor. C'est lui qui bénéficie, dans le temps du cas d'usage, des ressources de l'équipe *Big Data*, des budgets informatiques associés, qui les oriente et qui décide.

Nous sommes trop peureux pour supporter les responsabilités.
G. K. Chesterton

Au bout de quelques mois, lorsqu'il s'agira de faire le bilan *business* de l'initiative, ce sera au sponsor de s'y atteler, et il en sera tenu responsable. Il porte donc, dans ses objectifs annuels, la réalisation des gains *business* issus de la data. Et donnera donc tout l'appui nécessaire à la démarche… parce que c'est son intérêt !

ILS NE L'ONT PAS FAIT…

En 2012, François Hollande est élu à l'Elysée. Ses premières décisions laissent entendre qu'une transformation digitale de l'État se prépare. En octobre 2012 est créé le Secrétariat Général pour la Modernisation de l'Action Publique (SGMAP), à laquelle est adjointe la structure Etalab, en charge de l'ouverture des data publiques lancée sous la précédente mandature.

Les ingrédients de départ, esquissés dans le chapitre « Libérer vos énergies » (p. 79-96), sont les bons : la création d'une équipe transverse, concentrant les expertises avancées et directement rattachée au Premier Ministre.

■■■

Copyright © 2018 Dunod.

Plusieurs années plus tard, quel bilan en tirer ? La révolution attendue vers l'État-Plateforme (plus de détails dans la partie « La data, pour faire autrement ! », p. 3) n'est pas arrivée, loin s'en faut.

Pour faire bouger les lignes en profondeur, il a manqué aux équipes en charge un mandat clair, et une autorité sur les autres ministères. Il faut lire entre les lignes d'Henri Verdier, dans son rapport de l'administrateur général des données, paru en 2016. Il y déplore les freins au mouvement : la prégnance d'une culture administrative n'encourageant pas le partage des data, une gestion budgétaire qui paralyse la circulation de l'information… *a fortiori* dans l'univers conservateur de l'administration, il est indispensable que le pouvoir politique pèse de tout son poids, pour casser les silos et libérer les énergies.[1]

1. www.modernisation.gouv.fr/sites/default/files/rapport_agd_decembre2015.pdf.

© Dunod - Toute reproduction non autorisée est un délit.
Copyright © 2018 Dunod.

Copyright © 2018 Dunod.

5 CHAPITRE Réussir votre exécution

Les PowerPoint sont les aliments les plus consommés sur le marché. Les consultants et les entreprises les produisent à haute fréquence, en série. Souvent du même bois, d'ailleurs, sur le fond et la forme. Sur des slides, les stratégies 2020 des entreprises se ressemblent. Jusqu'à donner l'impression, dans certains cas, de vilains et maladroits « copier-coller ». En termes d'ambition data, le mimétisme est roi. Votre différenciation, ce sera surtout votre capacité à exécuter. Et donc votre vélocité.

Copyright © 2018 Dunod.

Copyright © 2018 Dunod.

QUEL AFFREUX JARGON !

VÉLOCITÉ

Concept clé, moins réducteur que celui de productivité, qui porte deux notions essentielles à la bonne marche collective : la rapidité, mais aussi la direction. Être véloce, c'est aller vite, et dans la bonne direction.

Les pages qui suivent, sous forme de neuf enseignements, vous aideront à réussir le « comment », en évitant (tant que possible !) les écueils classiques.

1 | LE DERNIER MÈTRE, TU RÉUSSIRAS !

La data reste abstraite. Des chiffres, des équations, des scores. Cela peut donc rapidement devenir froid et austère pour ceux qui ne baignent pas dedans tous les jours. C'est souvent le cas des utilisateurs finaux, en bout de chaîne, ceux qui doivent recevoir la data, et effectuer certaines actions en conséquence. Le marché français compte à peu près six cent mille ingénieurs, sur vingt-huit millions d'actifs. Partez du principe qu'une grande partie de vos collaborateurs n'aiment pas particulièrement les chiffres.

Dès le démarrage de vos cas d'usage data, il faut penser à votre utilisateur final : *Qui est-il ? Quelle est son appétence aux chiffres ? Quel temps va-t-il consacrer à la lecture et la compréhension des chiffres ou indicateurs que vous allez lui communiquer ? Quelles actions attendez-vous de lui, en retour ?* Il est crucial d'apporter des réponses à ces questions. Considérez-les comme des spécifications à part entière du projet. Et ceci est valable pour toutes les activités. Un nouvel algorithme pour détecter des anomalies de fabrication en usine ? Qui va recevoir l'information, quels sont ses leviers pour corriger le problème ?

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

Une détection plus efficace des fraudeurs parmi vos joueurs en ligne : comment agir, en sachant qu'il s'agit de probabilité, et que certains sont suspectés à tort ? Le souci du dernier mètre, c'est le souci de l'utilisateur final. Malheureusement, la data et l'utilisateur final, trop souvent, ne se trouvent pas. Voici quelques conditions importantes à remplir, pour éviter les rendez-vous manqués.

- **Être compris.**
  Un indicateur n'est jamais aussi simple qu'il en a l'air. Faites le test : dans une présentation chiffrée, arrêtez-vous, d'un coup, sur un indicateur exprimé en pourcentage, et posez la question : *Pouvez-vous me l'expliquer dans une phrase simple ?* Jugez le résultat. À vrai dire, la meilleure façon pour un chiffre d'être compris, c'est de disparaître… en tant que chiffre. De devenir autre chose : une couleur, un logo, une flèche… pensez à l'exemple de la météo. Des volumes colossaux de calculs, de data, pour former, à la fin, sur l'écran de votre application météo, un simple nuage, un soleil ou un brouillard.
  Il est rare d'avoir besoin de garder un chiffre absolu. Et il est plutôt utile d'essayer de le faire disparaître, pour les utilisateurs finaux. Pour déclencher une alerte par exemple, fixez le seuil et affichez un pictogramme rouge lorsqu'il est dépassé.
  En bref : pour bien communiquer des chiffres, évitez les chiffres !
- **Être adopté.**
  Il ne suffit pas qu'un chiffre soit compris pour qu'il soit utilisé et donne lieu à l'action efficace. Il faut que l'utilisateur ait envie de l'utiliser. La data doit ici être pensée sous l'angle de l'émotion. La question devient : *Qu'est-ce que vous souhaitez que l'utilisateur final ressente au contact de votre data ?* : sérénité, urgence, inquiétude… dans l'expression look and feel – *observer et ressentir*, n'oubliez pas le feel – le *ressentir*. L'interface en aval, qui portera la data jusqu'à l'utilisateur final dans son poste de travail, est souvent plus importante que la précision du calcul en amont. D'où la nécessité d'avoir recours, dans votre équipe *Big Data*, aux *designers* (comme expliqué dans la partie « La diversité, avant tout ! », p. 85-92).

Copyright © 2018 Dunod.

- **Être utile !**
  Vous souriez, mais de nombreux projets data échouent simplement parce que la data n'a pas été communiquée au bon utilisateur final, celui qui a le levier d'action. Que la data s'est perdue, entre les mains d'intermédiaires. En data comme au football, allez droit au but ! Veillez donc, au démarrage de chaque cas d'usage, à bien formaliser l'ensemble des actions à déclencher en aval.
  Et vérifiez bien que votre utilisateur final soit celui qui a les leviers pour les mener.

QUEL AFFREUX JARGON !

PRE-MORTEM

Beaucoup d'entreprises se gargarisent de faire des *post-mortem*. Ou de fêter les erreurs. C'est très bien, et loin de nous l'idée de les décourager. Le droit à l'erreur est un droit imprescriptible, et dont il faut pouvoir user, sans subir l'opprobre collectif. Mais convenez qu'il est mieux d'éviter les erreurs. Préférez-lui l'exercice du *pre-mortem*, car il permet d'anticiper. Qu'est-ce exactement ? Au démarrage d'un cas d'usage data, déclenchez une réunion avec les principaux acteurs du projet, dont le fameux sponsor, et faites-en un exercice de pensée : *Vous êtes six mois après la fin du projet, et ce dernier a été un désastre. Vous avez quinze minutes pour débriefer !* Pour sophistiquer l'exercice, ajoutez : *En termes de data, le projet avait été un succès. Beau score, belle prédiction. Le souci : personne dans l'entreprise ne s'en est servi !*
Dans la série « changer un des deux mots d'une pratique managériale connue, pour la booster », intéressez-vous au *feed-forward*, moins connu, mais plus utile que le fameux *feed-back*.

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

UNE ASTUCE !

Écouter l'utilisateur final de la data est une clé pour réussir. Lorsque vous aurez livré l'outil qu'il utilisera, intégrez un bouton qui permettra à l'utilisateur d'exprimer un « j'aime » ou « j'aime pas ». Ce bouton aura deux vertus : embarquer les utilisateurs dans une logique précieuse de co-construction, mais aussi (et surtout !) améliorer le modèle statistique en tant que tel.
Tout en améliorant le niveau de labellisation (la définition dans la partie « Humain, trop humain ! », p. 74-78). Prenez l'exemple d'un algorithme qui prédirait des moments de vie client (déménagements, changements de véhicule, départ à la retraite), pour des conseillers bancaires. Il serait utile que le conseiller bancaire puisse indiquer, sur son poste de travail : *Mais non, mon client ne va pas déménager, il attend juste un enfant, rien de plus.*

Copyright © 2018 Dunod.

2 DE LA RÉGLEMENTATION, TU TE PRÉOCCUPERAS !

Vous êtes tenus d'appliquer la réglementation européenne GDPR à partir de mai 2018 (déjà évoquée dans la partie « La data, pour ne pas disparaître ! », p. 23-30). Elle représente une rupture fondamentale dans le droit européen relatif aux data et creuse l'écart qui le sépare du droit américain. Un vrai casse-tête (onéreux) à venir pour les entreprises qui opèrent à l'international. Le GDPR impose de nombreux principes, dont vous retrouverez les principaux :

- Vous devez nommer un Data Protection Officer (DPO) dans votre organisation.
- Vous devez obtenir le consentement explicite de la part de vos clients pour utiliser leurs data, au moment de la collecte.
- La finalité, exprimée au moment de ce consentement, doit être en ligne avec votre utilisation effective de la data.
- Vous ne devez collecter que les seules data indispensables à l'exercice de cette finalité.
- Les data doivent être sécurisées, et susceptibles d'être auditées.
- Vos clients, à tous moments, ont le droit de consulter les data collectées les concernant, et de se voir expliquer les utilisations qui en sont faites, ainsi que les calculs effectués.
- Vos clients, à tous moments, ont le droit d'exiger que les data les concernant soient effacées.

Pour finir, le régulateur a une nouvelle importante à vous annoncer. Et faites-lui confiance pour veiller à ce que vous l'ayez bien digérée : les data de vos clients ne sont plus à vous. Nous en voulons pour preuve le fameux article vingt, du même règlement : « … *Le client a le droit d'avoir ses data, à sa demande, transférées d'un acteur A à un autre acteur B…* » Qu'est-ce à dire ? Qu'à la condition d'avoir l'autorisation de son client (pas le plus difficile, à portée d'un clic !), Google pourra demander à BNP Paribas l'ensemble de ses transactions bancaires ; Amazon pourra exiger

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

d'EDF le détail de sa consommation d'énergie des douze derniers mois ; le fabricant de matelas Matelsom pourra forcer Withings à lui fournir les data relatives à la qualité de sommeil de ses clients.

À tous moments, d'autres acteurs économiques (à commencer par vos concurrents) peuvent pénétrer chez vous, et, d'autorité, consulter ou faire évoluer les data du client en votre possession. C'est un peu l'extension du mandat de perquisition à la sphère économique.

À noter que cet article vingt va dans le sens de l'histoire, décrit au fil des pages de cet ouvrage, qui tend à toujours plus d'ouverture, de partage de la data, entre les acteurs économiques et publics.

Attention ! Les peines encourues en cas de manquement sont lourdes : des amendes jusqu'à vingt millions d'euros ou 4 % du chiffre d'affaires de votre entreprise (si votre chiffre d'affaires représente plus de vingt millions d'euros).

Structurez le plan d'action que vous allez concevoir, en réponse à ce règlement, autour de la différence fondamentale entre les mondes de la data analytique et de la data transactionnelle. L'interprétation du jeu de contraintes réglementaires peut évoluer entre les deux mondes. Pour en savoir plus, rendez-vous à la partie « Le Grand Soir technologique, c'est ce soir ! » (p. 61-68).

Enfin, saisissez la réglementation comme une opportunité. Si elle implique une couche supplémentaire de contrôle, et donc de bureaucratie, elle ne fait que formaliser des principes sains, qui devraient déjà guider votre activité *business*. Comme la préoccupation de bâtir un lien de confiance avec vos clients, en protégeant leurs data et leurs intérêts. Pour démarrer votre aventure data, et lancer vos premiers investissements financiers, arguer de la régulation, et sensibiliser au danger de l'immobilisme peut d'ailleurs devenir votre carte maîtresse ! Ne pas prendre le sujet au sérieux, c'est surtout se mettre entre les mains des *hackers*.

Concluons par quelques chiffres (un peu inquiétants) sur la maturité de nos PME sur le marché français. Car la réglementation data ne vaut

Copyright © 2018 Dunod.

pas (et de loin !) que pour les grandes entreprises du CAC40 : 32 % des PME ont été attaquées ces douze derniers mois ; 51 % se disent préoccupées par le danger ; 63 % se sentent peu ou pas préparées ; et seulement 10 % d'entre elles ont un expert en cybersécurité (à comparer à 55 % pour les entreprises de plus de mille collaborateurs).

ILS L'ONT FAIT !

L'exemple de la banque Goldman Sachs, déjà citée dans la partie « La data, pour faire autrement ! » (p. 13-22), est particulièrement instructif. En résumé : elle s'est lancée dans le développement d'une plateforme technologique automatisant un nombre colossal et croissant de transactions financières. Un volume inouï de data, donnant lieu, de façon continue, aux calculs d'algorithmes parmi les plus sophistiqués du marché.

Dans ce contexte, *quid* de la capacité de contrôle du régulateur ? À première vue, elle s'amenuise, à mesure que les algorithmes de Goldman Sachs se sophistiquent. Mais le banquier a pris le sujet dans l'autre sens : ils ont conçu la fameuse plateforme, dès le début, avec l'objectif de répondre aux enjeux de régulation. Qu'est-ce à dire concrètement ? À chaque fois que les algorithmes s'apprêtent à avoir un impact sur la vraie vie, donc à effectuer de véritables transactions, ils doivent pouvoir être compris *via* une documentation transmise à une autorité de contrôle externe.

La plateforme fait une découpe claire entre le monde de l'analytique, dont elle assume en quelque sorte de faire une boîte noire, et le monde du transactionnel, soumis à un contrôle extrêmement étroit.

Un humble avis : cette découpe est une merveille conceptuelle, mais dure à tenir d'un point de vue opérationnel. L'exemple n'en reste pas moins intéressant, car il montre la nécessité, dans tous vos cas d'usage data, de faire en sorte que les data puissent être auditées par le régulateur, et ce dès leur conception.

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

3 | LES ALGORITHMES, TU NE SOPHISTIQUERAS PAS !

> **Il faut prendre une statistique sur deux avec des pincettes.**
>
> **Sylvain Tesson**

Le sociologue Jean Baudrillard évoque souvent notre plaisir, un peu morbide, à voir les nombres croître indéfiniment. Vos *Data Scientists* sont souvent atteints du même mal : faire progresser la performance d'un score statistique, coûte que coûte, sans s'arrêter. Le souci, c'est que votre entreprise n'est pas un institut académique, qui gagnerait sa vie (et sa réputation) à publier des articles de recherche. Ne vous trompez pas : la performance d'un algorithme est importante, essentielle même. Ce propos est différent : à partir d'un certain niveau de performance, la progression statistique incrémentale, trouvée par vos statisticiens, se fera au détriment de la capacité de l'algorithme à être opérationnalisé, dans la réalité.

Qu'est-ce à dire concrètement ? Prenez un exemple : vous cherchez à prédire la réussite à six mois des candidats fraîchement recrutés sur votre plateforme téléphonique. Un *Data Scientist* que vous allez faire plancher sur la problématique vous demandera, pour commencer son analyse, l'ensemble des data à disposition sur les candidats passés. Les data présentes sur les CV (âges, parcours académiques, entreprises précédentes…), et éventuellement celles issues des retours qualitatifs des recruteurs, suite aux entretiens (remarques, commentaires libres, scores). Le *Data Scientist* produira un premier score de prédiction, que nous supposerons efficace. Cette prédiction s'ajuste parfaitement à la vraie vie. Pourquoi ? Tout simplement parce que les recruteurs pourront l'appliquer aux futurs candidats.

L'algorithme se nourrit en effet de data qu'ils peuvent collecter dans leur quotidien (CV et retours qualitatifs suite à leurs entretiens). Vous pourriez (et devriez !) vous arrêter là, et consacrez l'énergie de vos équipes à réussir l'intégration du score dans le poste de travail de vos recruteurs. Mais le *Data Scientist* voudra continuer : *Laissez-moi faire,*

Copyright © 2018 Dunod.

et je vais vous dénicher d'autres variables, encore plus explicatives. Et ce dernier d'aller regarder le nombre d'amis sur Facebook des meilleurs recrues, de leur poser des questions sur leurs goûts sportifs, littéraires, etc., puis d'intégrer leurs réponses à son modèle. Certes, il pourra en améliorer la performance, mais il l'éloigne de la réalité.

Dans la réalité, imaginez-vous vos recruteurs interroger les candidats sur leurs goûts sportifs ? Aller vérifier leur nombre d'amis sur Facebook ?

Dans la seconde phase de son travail, le *Data Scientist* a perdu le réel. Faut-il le censurer ? Non ! Car certaines des variables supplémentaires qu'il aura détectées pourront être intégrables dans la réalité. Pour reprendre notre exemple : si votre *Data Scientist* trouve que la durée de trajet entre le lieu de travail et d'habitation est une variable fortement explicative de la qualité d'un candidat, rien n'empêche vos recruteurs de rajouter systématiquement cette question lors de leurs entretiens !

> **Pour être valable, toute théorie, quelle qu'elle soit, doit être confirmée, tant dans ses hypothèses que dans ses conséquences, par les données de l'observation.**
>
> **Maurice Allais**

Tirez-en un principal enseignement : n'intégrez dans votre modèle statistique que des variables susceptibles d'être disponibles dans la réalité, juste avant que l'algorithme ne produise son résultat. Traduit en langage plus statistique, il s'agit de trouver un équilibre. Plus votre modèle est sophistiqué, plus il gagnera en précision (efficacité), mais plus il perdra en couverture (volume).

Avant chaque cas d'usage data, posez-vous la question du bon équilibre. La priorisation de la couverture sur la précision déplaira certes à vos *Data Scientists*, mais vous dégagerez souvent plus de valeur, et plus rapidement, de leurs travaux.

Ayez donc le souci de construire vos algorithmes avec les data disponibles, non les data potentiellement accessibles (ou fantasmées comme telles). La captation d'une data supplémentaire a toujours un coût. S'il n'est pas financier, il est opérationnel. Pesez bien, avec votre *Data Scientist*, l'incré-

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

ment attendu en termes de performance du modèle. *D'accord pour intégrer cette nouvelle data, mais que m'apportera-t-elle en retour ?* Peu de data, mais de qualité, valent mieux que beaucoup de data de qualité médiocre. Itérez !

UNE ASTUCE !

En data comme en médecine, l'effet placebo peut être utile. Dans certains cas d'usage data « humain » (rendez-vous à la partie « Humain, trop humain ! », p. 74-78), vous avez peut-être déjà noté que la data et les prédictions d'algorithmes n'ont pas beaucoup de valeurs intrinsèques. Ils servent surtout de détonateurs, pour des actions commerciales par exemple. En disant à un conseiller bancaire *Ce client va déménager, appelle-le*, vous souhaitez surtout le motiver pour décrocher son téléphone, appeler son client et faire des affaires. Et *in fine*, soyons francs, dans le taux de transformation commerciale de l'appel téléphonique, la prédiction du moment de vie par l'algorithme compte assez peu. Ce qui fait la vente, c'est souvent le simple fait d'appeler le client ! D'où la tentation de l'effet placebo, c'est-à-dire de mettre vos utilisateurs finaux en contact avec des alertes, des prédictions aléatoires, non assises sur des algorithmes véritables. La vertu ? Pas le fait de jouer sur leur crédulité bien sûr, mais de tester à blanc, avant d'investir davantage, les procédures opérationnelles en aval (dont l'importance est rappelée dans la partie « Le dernier mètre, tu réussiras ! », p. 99-102). Cette astuce est comme une démonstration par l'absurde : ce qui compte dans un projet *Big Data*, c'est, davantage que le niveau de sophistication de l'algorithme, la maîtrise opérationnelle de bout-en-bout !

Copyright © 2018 Dunod.

4 | UN PROJET INFORMATICO-INFORMATIQUE, TU ÉVITERAS !

Nous ne le répéterons jamais assez : pour valoriser la data, il vous faut mettre le *business* aux commandes, en responsabilité ; les fameux sponsors mentionnés (dans la partie « Un mandat clair, sinon rien ! », p. 93-96). Votre aventure data ne doit pas se résumer à une commande donnée à vos équipes informatiques pour la construction d'un *datalake*. Le pire ? C'est que vous risquez d'y arriver, en y mettant ce qu'il faut d'investissements financiers. Vous aurez votre beau *datalake*, flambant neuf. Mais il restera là, trônant, tel un roi sans royaume, car il ne sera pas raccroché à vos outils opérationnels, en aval.

Ce sont précisément ces raccrochements qui, en consommant ces data, apporteront le retour sur investissement à toute l'opération. La data ne doit pas être un projet informatique ; pas plus que votre stratégie d'entreprise est une stratégie informatique. La technologie doit permettre sa bonne déclinaison opérationnelle, rien de plus. Deux raisons principales pour finir de vous convaincre.

- **Les data sont des objets « métiers » par excellence.**

La valeur d'un algorithme est moins liée à la sophistication de son modèle mathématique qu'à la qualité des data intégrées dans le modèle. Or, rien de plus fragile que la qualité d'une data. Elle peut disparaître à la moindre rupture de la « chaîne du froid », pour ainsi dire. Entre sa collecte primaire et son usage final, il faut un expert, un sachant, pour évaluer sa qualité, à chacune des étapes. Sans ça, c'est l'efficacité globale du cas d'usage data qui est mise à mal.

- **Les data ne produisent des résultats *business* qu'à condition d'être utilisées.**

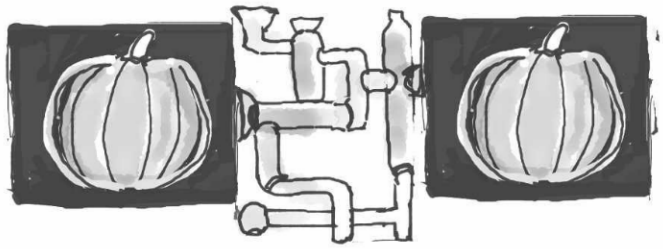
Nous touchons presque à la tautologie, du type *Tous les gagnants au LOTO avaient joué au LOTO*. C'est une évidence, les cas d'usage, « humain » comme « machine », doivent conditionner jusqu'à l'architec-

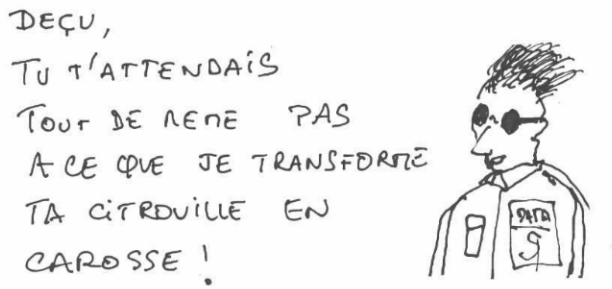
Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

ture du *datalake* (évoquée dans la partie « Le Grand Soir technologique, c'est ce soir ! », p. 61-68). D'autant plus que le facteur clé de succès tient essentiellement à la fluidité de vos procédures opérationnelles en aval.

À NOTER !

Dans votre entreprise, il y a de fortes probabilités que des outils de type *Big Data* soient déjà utilisés par vos équipes, sans qu'ils n'aient l'apparence d'un *datalake* en tant que tel. Il s'agit des outils digitaux (achats publicitaires, gestion des sites Web…). Ils sont d'une constitution bien différente de vos systèmes informatiques traditionnels. Ils brassent un fort volume de data non structurées, sont sur le *cloud* (à opposer à un hébergement des data en interne) et permettent des activations en temps réel (comme l'achat de telle ou telle requête sur Google). Point intéressant : le choix et l'adoption de ces outils s'effectuent plutôt du côté des équipes marketing. Si vos équipes informatiques ne sont pas encore très à l'aise avec la matière *Big Data*, conseillez-leur de regarder sous le capot de ces outils digitaux. Ça tiendra lieu de formation accélérée !





Copyright © 2018 Dunod.

5 | SUR LES DATA EXTERNES, TU NE FANTASMERAS PAS !

Le *Big Data*, c'est d'abord l'émergence de trois nouvelles sources de data. Outre les data digitales et comportementales, il y a les data externes. C'est aussi celles qui suscitent (étrangement) le plus d'attentes : *Si nous pouvions attraper à l'extérieur telle ou telle data, nous pourrions faire telle ou telle chose…* Sachez-le : cela tient, quelques fois, de l'incantation. En termes de data externes, le marché français n'est pas encore mature. Il faudra encore plusieurs années pour qu'il le devienne. Ni les systèmes informatiques, ni d'ailleurs les organisations dans leur ensemble, ne sont prêts à ouvrir largement leurs data sur l'externe, pour les échanger.

Faisons le tour des sources de data externes actuellement disponibles.

LES ACTEURS PUBLICS

Si la France n'est pas le pays occidental le moins avancé (loin de là !), encore très peu de data sont rendues accessibles par les organismes publics. Rendez-vous sur www.data.gouv.fr pour les consulter (… et dans la partie « La data, pour mieux faire ! », p. 3-12).

LES ACTEURS PRIVÉS

Si beaucoup d'entreprises se sont (enfin !) décidées à développer des API sur leurs systèmes informatiques pour rendre consommables leurs data par des acteurs externes, les exemples concrets d'échanges sont encore très rares en France. Au-delà des raisons technologiques, il y a des explications plus culturelles. Pour échanger des data avec des partenaires externes (des futurs déménageurs pour EDF ; des probables

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

pépins techniques pour les constructeurs automobile…), encore faut-il leur donner une valeur, sonnante et trébuchante. Comment s'y prendre ? Le troc ou la vente ?

En l'absence de marché confrontant une offre et une demande, avec un nombre minimum d'acteurs réunis, bien difficile de convenir d'une échelle de prix partagée. Il existe bien quelques *consortium* entre acteurs privés pour partager leurs data. Nous avons déjà cité l'exemple des médias français (dans la partie « La data, pour faire autrement ! », p. 13-22). Mais ils sont bien rares ! À noter qu'une disposition de la Loi pour la République Numérique d'Axelle Lemaire va obliger certains acteurs privés à communiquer un certain nombre de data dites d'intérêt général. La loi établit trois catégories : les data issues d'une délégation de service public (par exemple, une municipalité qui délègue à une entreprise privée la gestion de ses transports, pourra exiger, en retour, les data associées) ; les data liées aux conventions de subventions associatives, pour celles supérieures à 23 000 euros ; enfin, les data qui pourraient être nécessaires à des enquêtes INSEE, à des seules fins de statistiques (immobilier, énergie…).

LES ACTEURS DE LA PUBLICITÉ DIGITALE

Ce sont certes des acteurs privés, mais considérez-les dans une catégorie à part. En effet, leur spécificité est qu'ils vivent directement de la commercialisation de leurs data, à des fins publicitaires. Nous parlions, précédemment, des difficultés à attribuer une valeur à la data ? Eux n'ont pas le même souci. La publicité digitale fonctionne en effet sur la logique des marchés financiers, avec des enchères automatisées. Tout est codé, partagé, standardisé entre les acteurs. Google sait exactement ce que vaut la requête « crédit simulation » sur son moteur de recherche. Facebook la valeur d'un profil « avocat parisien célibataire amateur de bons vins » sur son réseau social. Ces data vous sont-elles accessibles ? Oui, bien sûr ! Mais en mettant deux conditions à l'échange, encore bien contraignantes.

Copyright © 2018 Dunod.

D'abord, il vous faudra les acheter (plutôt justifié). Puis, vous ne pourrez consommer ces data achetées que sur leurs seuls espaces publicitaires (moins justifié). Qu'est-ce à dire ? Prenons un exemple. Vous vendez des abonnements téléphoniques et vous souhaitez identifier, avant qu'ils ne vous quittent pour vos concurrents, vos clients les plus à risque. Google peut vous aider. *Via* la fonctionnalité de ciblage déjà présentée (dans la partie « La data, un patrimoine vivant ! », p. 48-55), vous pouvez atteindre ceux de vos clients qui tapent les requêtes « abonnement » ou « forfait » sur le moteur de recherche Google.

Mais pour les identifier, et faire le lien avec votre base de clients non anonymisés (avec un nom et un prénom précis), il vous faudra acheter des liens sponsorisés publicitaires sur Google… et encore, vous ne connaîtrez pas leur identité précise, à moins qu'ils n'aient cliqué sur votre lien publicitaire et ne soient arrivés sur votre site.

Pour résumer : ces acteurs digitaux savent beaucoup de choses, donnent peu, et à un tarif élevé. Des partenaires certes… mais un peu particuliers !

Les data externes ne sont donc pas la panacée. Pas encore, mais elles le seront demain, lorsque les acteurs privés (et pas seulement les régies publicitaires) auront fini leur mue vers plus d'ouverture externe. Ce manque de maturité collective, sur le marché français, est aussi une opportunité : soyez à l'affût des quelques data disponibles, car il y a encore peu d'acheteurs intéressés et compétents !

QUEL AFFREUX JARGON !

FICHIER CENTRAL DES AUTOMOBILES (FCA)

Le FCA est un fichier informatisé placé sous la responsabilité du ministère de l'Écologie, du Développement et de l'Aménagement Durables français. Sa gestion a été déléguée à l'Association Auxiliaire de l'Automobile (AAA).

■■■

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

■■■

Il contient l'ensemble des immatriculations françaises, le parc de véhicules associés, ainsi que des informations nominatives sur les conducteurs.
Ce fichier est intéressant car il préfigure le marché mature à venir. Au début, un fichier semi-public, mis à disposition du marché. Et un lot de revendeurs, des *brokers*, qui s'emparent du fichier, le reconditionnent, l'enrichissent d'autres data, pour le revendre, à un prix plus élevé. Face à la multiplicité de ces intermédiaires d'un nouveau type, il est encore compliqué d'apprécier, *a priori*, le rapport qualité/prix. Il vous faudra malheureusement tester, au cas par cas !

En résumé, concentrez-vous d'abord sur vos gisements de data internes, avant d'entreprendre la grande tournée des data externes disponibles sur le marché.

Copyright © 2018 Dunod.

6 | LES DATA DIGITALES, TU CHÉRIRAS !

À chaque fois que vous cliquez sur votre souris, tapez sur votre clavier ou *swipez* (*swipez* ? rendez-vous à la partie « Tes clients, tu associeras ! », p. 118-122 pour la définition), vous laissez des empreintes. Le digital permet de collecter une quantité formidable de data sur les comportements, les attentes et les profils de vos clients comme des prospects. Imaginez une caméra à l'épaule de chaque client dans un supermarché : connaître le rayon devant lequel le client rêvasse ; les deux produits entre lesquels il hésite ; le produit qu'il prend dans les mains puis repose ; le soupir qu'il pousse à la vue de la queue trop longue à la caisse, devant laquelle finalement il renonce…

Votre organisation collecte déjà les data digitales de vos internautes par le biais d'outils de webanalyse. Pourtant, il est peu probable qu'elle les mette en action, au service de meilleurs parcours ou de meilleures offres. La discipline est encore jeune, mais les premières organisations (largement anglo-saxonnes) à avoir mis la data digitale au centre de leurs stratégies, en récoltent déjà les fruits. Pensez à Google, mais aussi à Netflix, Amazon, Tesco ou Booking… Autant d'entreprises au sein desquelles les notions d'agilité et d'expérimentation continue sont portées et diffusées par le *management*, introduites dans les procédures internes et mises au service direct du *business*.

Et vous, où en êtes-vous en termes de data digitales ? Autant les data externes peuvent attendre, autant avec vos data digitales, il faut avancer ! Et vite. De quelles data parlons-nous ? Des navigations digitales de vos clients et prospects sur votre écosystème de sites et d'applications mobiles.

Ces data représentent un gisement souvent insoupçonné par les entreprises. Pourtant, elles apportent des signaux d'information inégalés pour compléter la connaissance de vos clients et enrichir l'ensemble de vos leviers d'activation (digitaux, mais aussi de marketing direct). Ces data digitales présentent également une caractéristique essentielle

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

par rapport à d'autres sources de data : elles sont collectées de façon discrète, c'est-à-dire indolore pour l'utilisateur. Contrairement à des études en laboratoire ou des panels (méthodes évoquées au chapitre « La collecte pour tous ! »), la collecte des data n'influence pas le comportement des personnes observées. Le comportement étudié est donc dit « naturel ».

Si beaucoup d'entreprises passent à côté des data digitales, ce n'est pas par instinct autodestructeur, rassurez-vous. C'est parce que le digital bouleverse fondamentalement la façon d'identifier vos clients. Derrière l'écran, plus de noms et de prénoms. Le *cookie* devient le passeport. Qu'est-ce qu'un *cookie* ? C'est une suite de caractères alphanumériques, non intelligible pour le commun des mortels, qui se place dans votre navigateur web de façon durable. Leur durée de vie peuvent aller de quelques secondes à deux ans, et ils permettent, entre autres, aux sites de vous reconnaître lors de vos visites successives et d'améliorer la fluidité de vos navigations (mémorisations de votre mot de passe, personnalisations du message d'accueil, suggestions de produits…). Cependant, peu d'entreprises maîtrisent la notion de *cookie*, ainsi que le volume colossal de data à stocker et traiter pour le valoriser. Collecter les data digitales, c'est entrer de plein fouet dans l'ère du *Big Data*. D'où les difficultés et le retard de certains acteurs…

COMMENT ÇA FONCTIONNE, CONCRÈTEMENT ?

Comment poser un *cookie* à vos clients, de façon à le reconnaître en ligne ? Pour cookifier vos internautes, c'est simple : il vous faut trouver des moments digitaux pour lesquels vous êtes sûrs de l'identité nominative de l'internaute intercepté. Il y a plusieurs façons de procéder. D'abord par vous-mêmes : si vous avez un espace-client digital, sur lequel vos clients s'identifient *via* un mot de passe, vous pouvez cookifier vos clients avec confiance, juste après leur authentification. À chaque fois que vos clients reçoivent l'un de vos e-mails, vous pouvez également leur poser un *cookie*, parce que vous connaissez

■■■

Copyright © 2018 Dunod.

l'adresse à laquelle vous avez envoyé le message. Ces deux premiers leviers de cookification sont efficaces, mais pas forcément suffisants pour identifier tous vos clients. Pour parfaire l'opération, activez un troisième levier, en demandant à des partenaires externes de cookifier vos clients à votre place. Prenons un exemple concret : vos clients renseignent leur e-mail sur de nombreux sites chaque mois, pour consulter une annonce immobilière, pour activer un espace-client, pour s'inscrire à une newsletter… Rapprochez-vous de ces sites et proposez-leur de les rémunérer, pour qu'à chaque fois qu'un de vos clients renseigne son e-mail sur l'un de leurs formulaires, il lui pose votre cookie. Pour ce faire, il suffit de lui transmettre votre base d'e-mails !

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

7 | TES CLIENTS, TU ASSOCIERAS !

Nous l'avons évoqué dans la partie « De la réglementation, tu te préoccuperas ! » (p. 103-105). Le mouvement massif de digitalisation des activités économiques, comme des vies quotidiennes, a son revers : il porte une exigence forte de réciprocité de la part des individus. Les clients, les consommateurs, communiquent davantage de data aux acteurs qui les entourent. Ils souhaitent obtenir davantage en retour. C'est le fameux *value for data* (mentionné dans la partie « La data, pour ne pas disparaître ! », p. 23-30).

Prenons l'exemple du secteur financier : pendant des années, les banques ont collecté les data liées aux paiements de leurs clients. Beaucoup de clients, beaucoup de data, vous l'imaginez. Qu'en ont-elles fait ? Pas grand-chose, si l'on se place dans les chaussures d'un client. Désormais, ce dernier veut des services tangibles en retour : aide continue à la gestion de sa trésorerie, proposition d'épargne quotidienne, suggestions d'économies… Cette exigence de réciprocité va toucher tous les secteurs, dont le vôtre. Il faut donc s'y préparer !

Une projection : le marché actuel est loin d'être optimal dans la gestion de la data client. Autour d'un consommateur donné, imaginez des dizaines, voire des centaines d'entreprises différentes, qui ont en leur possession les mêmes data le concernant : numéros de téléphone, adresses postales, centres d'intérêt… chacune de ses entreprises ayant sa propre façon de collecter et structurer ces data. Imaginez, des centaines de fois répétés, les formulaires de saisie, les bases de data à construire, et le travail de compréhension à mener, alors que les data sont les mêmes, et que la personne la plus adaptée pour les comprendre, et les rafraîchir, c'est bien le client lui-même ! Est-ce bien optimal ?

Si nous prenions (le temps de ce paragraphe !) un angle libéral, le marché de la data client ressemblerait à quelque chose de cet ordre : des data client structurées, rafraîchies à un seul endroit, et sous la responsabilité de celui qui les génère, le client. Pour interagir avec un

Copyright © 2018 Dunod.

nouvelle entreprise, ce dernier n'aurait qu'à faire un clic pour communiquer ses data, plutôt que de remplir plusieurs champs de formulaire.

Pour ce faire, il faut que chacun d'entre nous, en tant que citoyen ou consommateur, bénéficie d'une sorte de passeport digital élargi, qui couvrirait des data plus riches que les seules data d'identité actuellement imprimées sur votre passeport physique. Une question se pose alors : Quelle solution permettrait d'y parvenir ? Deux extrêmes possibles, que nous allons caricaturer (ce ne sera pas la première fois).

L'APPROCHE ÉTATIQUE

C'est une piste à considérer sérieusement. L'État concevrait et mettrait à votre disposition l'outil rassemblant vos data personnelles. Il en a déjà beaucoup même si elles sont peu consolidées : pensez aux data liées à l'impôt, à la Sécurité sociale, aux transports. Pour autant, il existe deux freins forts à son développement : d'abord, la suspicion citoyenne. Les citoyens sont-ils prêts à faire de l'État l'arbitre de leurs consommations courantes ? Car dans cette configuration, ils le placeraient entre eux et les entreprises avec lesquelles ils interagissent (grande distribution, énergie, télécommunication, loisirs). Rien n'est moins sûr.

Autre frein (plus grave peut-être) : l'incapacité de l'État à livrer rapidement, puis à faire vivre dans le temps, ce type d'outil grand public. Il faut des talents, de la créativité et une attention à l'expérience client dont l'État est aujourd'hui malheureusement bien dépourvu.

Le précédent du projet de développement du Dossier Médical Personnalisé (DMP) est à ce titre édifiant. Il n'y a qu'à visiter le site sur ce qu'il en reste pour frissonner[1]. Exemple typique d'une initiative étatique pensée en chambre, sans aucun des ingrédients de succès de la nouvelle économie digitale : travail collaboratif et obsession de la convivialité des interfaces mises entre les mains des utilisateurs.

1. http://goo.gl/snvSaq

© Dunod - Toute reproduction non autorisée est un délit.
Copyright © 2018 Dunod.

Le prochain outil DMP qui rencontrera le succès et l'usage (car le besoin existe !), ce sont des acteurs privés qui le lanceront. Plusieurs en ont déjà commencé le développement…

L'APPROCHE PRIVÉE

C'est la piste la plus probable. Pensons aux 33 millions de profils Facebook créés en France. Que sont-ils, sinon les passeports digitaux les plus riches et le plus récents que chaque Français a à présenter ? Le réseau social californien ne s'y est pas trompé, qui vous propose, depuis 2008, d'utiliser votre identité Facebook (Facebook Connect) pour s'identifier sur d'autres sites, sans avoir à remplir de nouveaux formulaires. Aujourd'hui, *via* Facebook Connect, vous vous contentez de communiquer plus facilement à des sites tiers votre nom, prénom et mot de passe.

Demain, imaginez que votre banque, votre fournisseur Internet, vos programmes de fidélité, l'État, viennent s'alimenter aux data fraîches, et plus nombreuses, présentes sur votre profil Facebook : vos numéros, vos adresses, votre patrimoine. Depuis les messageries Messenger ou WhatsApp (toujours Facebook !), les marques parlent déjà à leurs clients, reconnus comme tels, *a priori*.

D'autres acteurs privés peuvent-ils encore prendre cette place privilégiée ? Pas sûr, tant la bataille sera rude, et beaucoup pense qu'elle est déjà perdue. Google, Facebook, Amazon ou Apple ont déjà collecté beaucoup de data, et pris une place de choix dans nos quotidiens. Une chose est sûre : l'acteur qui deviendra l'agrégateur de nos data personnelles sera celui qui aura rendu la collecte de ces data la plus naturelle, la plus indolore. Vous n'aimez pas l'effort.

Copyright © 2018 Dunod.

En résumé : le marché va tambours battants vers davantage de standardisation des data personnelles, et de centralisation à la main des particuliers. Demain, le mouvement touchera également les entreprises privées comme les institutions publiques.

Qu'est-ce que cela implique pour vous ? Que toutes vos actions data doivent respecter le principe suivant : ne faites rien, dans la collecte ou dans l'utilisation des data, que vous ne pourriez montrer, en toute transparence, à vos clients. Un peu à la façon dont on commence à dire, aux internautes de tous les jours : *Partez du principe que tout ce que vous faites en ligne est susceptible de devenir, un jour, public.* C'est un principe radical, et qui bouscule. À tous moments, vos clients doivent comprendre les data que vous collectez sur eux, et l'utilisation que vous en faites.

Conséquence logique : plutôt que de chercher à les récolter à leur insu, *via* des data externes par exemple, attachez-vous plutôt à enga-

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

ger vos clients, à leur poser directement les questions qui vous intéressent. Le tout est de le faire en réduisant leurs efforts. Donc plutôt que des formulaires administratifs et austères, usez d'interfaces fluides, ludiques et plaisantes : faites *swiper* vos clients !

QUEL AFFREUX JARGON !

SWIPER

Désigne le fait, pour les utilisateurs d'interfaces digitales, de glisser leur doigt de gauche à droite (ou l'inverse) sur l'écran, pour indiquer quelque chose, ou répondre à une question. Dans le cas de la célèbre application de rencontres Tinder, on *swipe* pour apprécier ou pas une personne. Ce principe d'ergonomie s'étend à d'autres interactions clients, dans lesquelles il faut répondre à une question fermée (oui/non). Le *swipe* remplace les formulaires et leurs champs de remplissage, pour lesquels l'utilisateur doit pianoter sur son clavier. Il peut permettre, en moins d'une seconde, à vos clients de dire « content/pas content », « j'ai déménagé/je n'ai pas déménagé », « je recommande/je ne recommande pas »… sans effort !

Copyright © 2018 Dunod.

8 DÉCRIRE ET PRÉDIRE, TU NE CONFONDRAS PAS !

Avant de démarrer votre analyse, de lancer vos statisticiens dans la conception de vos algorithmes, posez-vous cette question fondamentale : *Cherchons-nous à mieux comprendre le passé ou à anticiper le futur ?* La question paraît anodine, au premier abord. Moins, après y avoir réfléchi. Votre réponse est structurante. Elle conditionnera les méthodes statistiques choisies pour l'analyse, et jusqu'aux procédures opérationnelles activées en aval. Distinguez donc bien les deux démarches.

LA DÉMARCHE DESCRIPTIVE

Vous la connaissez sûrement sous le nom plus connu de *business intelligence*, dont la pratique est déjà très courante sur le marché. L'objectif est de tirer du sens du passé. D'en tirer des enseignements, à partir des vastes gisements de data existants. D'identifier des comportements répétitifs (appelés *patterns*), ainsi que des corrélations entre plusieurs variables, de prime abord indépendantes. Le passé doit être rendu compréhensible aux individus de chair et d'os, de façon à leur permettre de prendre des décisions. La méthode descriptive est celle qui « donne à voir ».

D'où le recours fréquent à la Dataviz (évoquée dans la partie « La data, ce n'est pas une arme ! », p. 37), et son lot de représentations graphiques, de tableaux et autres histogrammes colorés. Vous noterez, dans ces analyses, la présence d'indicateurs agrégés, comme des moyennes, des médianes ou des dispersions. Leur spécificité est de rendre compte d'un vaste ensemble de data, *via* quelques chiffres simples et peu de scores statistiques plus sophistiqués.

C'est la démarche descriptive qui vous permet aussi d'inférer des enseignements « macro », sur la base d'une analyse « micro ». C'est le principe de l'échantillonnage. Parce que vous avez observé telles

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

ou telles caractéristiques sur une population de taille restreinte, mais représentative de la population entière, vous pouvez les généraliser. C'est le principe (bien connu !) des sondages politiques.

En bref : la démarche descriptive vous révèle le passé, pour vous permettre (et non pour un algorithme) à vous d'éclairer le futur et de décider. Dans son ouvrage sur le suicide[1], le sociologue Durkheim montre la régularité du nombre de suicides chaque année, dans la société européenne au début du siècle dernier. Régularité qui permet d'éclairer les politiques publiques. Régularité descriptive, et non prédictive… Il n'a pas construit un modèle de probabilité de suicide pour chaque citoyen !

LA DÉMARCHE PRÉDICTIVE

Il s'agit là d'un tout autre raisonnement et modèle statistique. L'idée, en réaction immédiate à une nouvelle data, est d'être en capacité de prédire l'occurrence de tel ou tel événement.

Faisons l'analogie avec un médium. Vous lui posez une question sur l'avenir, il vous apporte une réponse, du tac-au-tac. Remplacez le medium par l'algorithme. Vous noterez que la prédiction, en tant que telle, est déjà le socle de beaucoup de vos procédures actuelles. Sur une plateforme téléphonique, elle porte le nom d'hypervision, métier qui consiste à prévoir les volumes d'appels, de façon à ajuster au mieux le volume de conseillers en réception. La maîtrise de cette discipline est d'ailleurs l'une des clés majeures de la compétitivité des centres d'appel.

Si la prédiction n'est pas une discipline nouvelle, elle change d'envergure, et de surface d'impact, avec le *Big Data*. Pourquoi ? L'explosion du volume des data intégrées dans le modèle, alliée à des modèles statistiques d'un nouvel ordre, auto-apprenants comme le *machine learning,* ont changé la donne. Les prédictions sont plus fiables, et s'ajustent

1. Émile Durkheim, *Le Suicide : Étude de sociologie*, 1897

Copyright © 2018 Dunod.

de mieux en mieux à la complexité du monde, en s'optimisant dans le temps. Demain, elles porteront les moteurs d'intelligence artificielle. Les applications de la méthode prédictive sont innombrables : depuis la détection des moments de vie, en marketing, jusqu'à la prédiction de fraude, en passant par l'obsolescence d'une chaîne d'assemblage industrielle, ou la gestion des ventes dans la grande distribution. En bref : cette démarche prédictive s'appuie certes sur le passé, mais a vocation, *via* un algorithme, à suggérer une décision précise, dans une situation future précise.

Plus votre activité sera automatisée, plus vos cas d'usage data seront « machine » (comme évoqué dans la partie « Humain, trop humain ! », p. 74-78), plus la méthode prédictive aura de sens, et apportera de la valeur. *Cherchons-nous à mieux comprendre le passé, ou à anticiper le futur ?* Notre question de départ revient finalement à une autre question, liée cette fois à l'activation de l'analyse, en aval : *Cherchons-nous à éclairer la décision d'un individu, ou à automatiser l'action d'une machine ?* Prenez le soin d'éclairer, au démarrage de chaque projet, l'angle souhaité. Nous avons vu trop de sponsors déçus, à la fin d'un projet, par des résultats descriptifs, alors qu'ils attendaient de l'actionnable précis, et immédiat.

> **Héroscope : prédiction flatteuse qui permet de fuir dans un avenir sublime la grisaille ou les pauvres tourments du quotidien.**
>
> **Alain Finkielkraut**

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

ILS L'ONT FAIT !

Le métier d'un assureur est précisément de prédire la probabilité du risque, pour un individu comme une collectivité. Par exemple, celle d'un accident automobile, d'un dégât des eaux ou encore d'une catastrophe naturelle. L'explosion récente du volume de data disponibles, notamment les data comportementales, ne laisse pas les assureurs indifférents. Pour eux, c'est une formidable opportunité de mieux affiner leurs modèles de prédiction.

Prenons l'exemple de la conduite automobile. Jusque-là, les assureurs, pour affiner votre prime, se contentaient des quelques informations captées lors de la souscription de votre contrat, *via* un formulaire classique : code postal, date du permis, antécédents… et ceci pendant toute la durée de votre contrat. Le suivi en temps réel de votre conduite, *via* un boîtier ou une application mobile, tel que le propose aujourd'hui de nombreux assureurs, permet de personnaliser davantage la prédiction de votre risque.

Votre dextérité de conducteur dans les tournants, le nombre de kilomètres effectués, votre temps moyen d'accélération en ville, sont des variables chaudes, à fort potentiel, jusque-là peu ou pas exploitées par les assureurs.

À noter : la prédiction de votre risque ne peut se réduire à la seule meilleure prise en compte de vos datas comportementales, pour une raison principale : les variables externes, qui ne dépendent pas de votre comportement, comme la météo, l'état des routes locales sur lesquelles vous roulez, etc. continuent d'expliquer une grande partie de votre risque. Vous l'aurez compris, les assureurs sont résolument du côté de l'approche prédictive, et non descriptive, du risque. Prédiction qui donne lieu, chaque jour, à la fixation de primes d'assurance différenciées, pour des milliers d'entre nous.

Copyright © 2018 Dunod.

IL TÉMOIGNE !

TRISTAN NEVEU

QU'EST CE QUI FAIT QUE L'IA (INTELLIGENCE ARTIFICIELLE) D'AUJOURD'HUI N'EST PAS CELLE D'HIER ?

Les récentes avancées de l'intelligence artificielle dans des domaines aussi variés que le langage ou la vision invitent clairement à l'enthousiasme. Un parallèle rapide avec les « hivers de l'IA » (dans les années 1970 à 1990, le financement et l'intérêt public pour l'intelligence artificielle se sont écroulés suite au manque de résultats pratiques) invite cependant à se poser la question de savoir en quoi la période actuelle diffère des échecs passés.

Tout d'abord, les résultats sont cette fois-ci au rendez-vous. L'utilisation de l'apprentissage profond, un des composants majeurs de l'IA, a permis de se rapprocher des capacités humaines dans certains cas (synthèse vocale, réponse à des questions écrites, systèmes de dialogue), de les égaler souvent (reconnaissance de types d'image, traduction automatique), voire de les dépasser (à titre d'exemple, AlphaGo est un système artificiel qui bat les meilleurs joueurs de go humains). Ces réalisations sont à opposer à la plupart des systèmes d'IA des années 1970, qui utilisaient souvent de simples règles de décision fonctionnant mal dans des usages non considérés par leur concepteur.

Parmi les facteurs qui expliquent cette révolution de l'apprentissage profond, on peut citer une très forte puissance de calcul disponible (utilisation des GPUs [processeurs graphiques] voire TPUs [processeurs de temps], calcul distribué), l'accès à un très grand nombre de jeux de données telles ImageNet, qui regroupe 14 millions d'images d'animaux et d'objets publiquement disponibles, et enfin la découverte de nouvelles techniques permettant l'entraînement de réseaux de neurones profonds. Ces trois facteurs ne sont apparus qu'assez récemment.

Par ailleurs, de nombreuses entreprises sont aujourd'hui soit fondées sur l'IA (Maluuba, ElementAI par exemple) soit dépendent activement de l'IA pour une partie de leurs opérations (Google, Facebook, Microsoft…). L'IA porte ses fruits et est activement incorporée dans des produits.

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

De plus, le développement de nouveaux outils liés au *Big Data* (notamment le *cloud computing*) et l'émergence de nouveaux comportements (le code et les papiers de recherche sont plus que jamais publics, les chercheurs redoublent d'efforts pour rendre toute expérience reproductible) font que désormais une entreprise désireuse d'incorporer de l'apprentissage statistique ou de l'IA à ses opérations peut le faire plutôt facilement, et à une échelle adéquate pour traiter des données internes souvent massives. À ce titre, on peut citer les activités de Criteo et d'AXA en France, ou encore l'utilisation massive de l'IA pendant la campagne présidentielle américaine. C'est à mon avis cette combinaison d'avancées rapides de la recherche et de démocratisation de l'utilisation de l'IA par les entreprises et les particuliers qui fait que la période actuelle peut être vue non plus comme une voie sans issue, mais comme le début d'une transformation en profondeur de nos usages.

Copyright © 2018 Dunod.

9 | DE LA DATA, TU TE MÉFIERAS !

Étrange conseil, dans un ouvrage dédié à la data ! Pensez au grand cru : pour apprécier le vin, il faut respecter une certaine modération. Il est aussi fortement indiqué de soigner le contexte : le verre, le plat associé, les camarades… La data, c'est la même chose.

La data n'est pas souhaitable, aimable, en tant que telle. Sa bonne valorisation *business* dépend de son intégration fluide dans votre environnement. À vous de lui fixer les garde-fous nécessaires. Un exemple simple pour s'en convaincre. Vous voici au volant, vers une destination qui vous est familière, sans pour autant que vous n'effectuiez le trajet tous les jours. Pour optimiser votre temps et éviter les embouteillages (vous êtes en ville), vous branchez le GPS. Après quelques minutes de trajet, le GPS vous suggère une route inhabituelle, puis une seconde route inhabituelle, qui vous éloigne du trajet classique. Question importante : à partir de quel moment allez-vous sérieusement arrêter de suivre le GPS aveuglément ? Cette question est cruciale. Parions qu'avec la vitesse exponentielle des progrès technologiques, le moment de ce réveil sera de plus en plus tardif, pire, que souvent il n'arrivera plus, ou trop tard.

À mesure que l'algorithme se perfectionne, la confiance que nous lui portons s'accroît naturellement, et notre vigilance, notre sens critique et jusqu'à notre responsabilité individuelle, diminuent d'autant. Face aux algorithmes qui nous entourent, ne perdez pas votre distance critique. Agissez en tant que chef d'entreprise, collaborateur, mais aussi citoyen. On peut citer trois raisons de ne pas placer une confiance aveugle dans les data.

LE RÉEL NE SE RÉDUIT PAS À DE LA DATA

Le courant de pensée « réductionniste » n'a jamais été aussi en vogue. Il désigne la tentation contemporaine à réduire la réalité à une suite continue de 1 et 0. Vos pensées, votre sommeil, vos actions, vos mobi-

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

lités et vos désirs… le mouvement actuel et massif de digitalisation participe activement de ce fantasme. Progressivement, il tisse un voile de data entre le monde et nous. En simplifiant, souvent à outrance, sa complexité.

Ce réductionnisme, nous nous y adonnons au quotidien, sans forcément nous en rendre compte. Un match de foot ne se réduit pas à un score ; un parcours de vie à quelques lignes d'un CV ; un profilage client à trois visites successives sur un site Web ; le succès d'une entreprise à son seul cours de Bourse ; la précarité à la seule inversion de la courbe du chômage. Trop souvent, la carte nous fait oublier le territoire.

LA MACHINE PEUT SE TROMPER

La remarque est assez prosaïque, certes. Mais il est essentiel d'intégrer, à tout moment, cette idée (pas très confortable !) : la machine est peut-être en train de *bugger*. Le GPS peut, sur certains trajets, se mettre à tourner en rond, vous faire passer par des chemins absurdes. Exercer au quotidien cet esprit de suspicion est inconfortable, car il oblige à rester en veille.

Le triste accident du vol Rio-Paris, en 2009, en témoigne, à l'envers : à l'origine, c'est une défaillance des sondes (probablement due au givrage), qui a soudainement stoppé le pilotage automatique de l'Airbus. Et le pilote a dû reprendre la main, en l'occurrence de façon malheureuse, réagissant davantage aux informations de son tableau de bord défectueux qu'à son intuition[1]. L'accident nous dit une chose essentielle : que le recours trop fréquent à la machine réduit progressivement l'autonomie humaine. Cette dépendance est d'autant plus préjudiciable en cas de défaillance de la machine.

L'individu, qui utilise un thermomètre, doit systématiquement, à la lecture du résultat, poser la question : fonctionne-t-il correctement ? Cette interrogation est la clé d'une complémentarité humains/machines

1. Voir l'ouvrage de Nicholas Carr, *The Glass Cage*, 2015.

Copyright © 2018 Dunod.

réussie. Vous n'êtes pas là pour répliquer de façon pavlovienne le fonctionnement des machines. Vous avez justement développé des machines pour cela ! Vous êtes là pour user de ce qui vous est propre, c'est-à-dire de votre sens critique. Il faut de la confiance certes, mais pas inconditionnelle.

> **Je ne crois qu'aux statistiques que j'ai moi-même trafiquées.**
>
> **Winston Churchill**

LES ALGORITHMES N'AIMENT PAS L'IMPRÉVU

Les algorithmes aboutissent à des prévisions suite à une analyse fine du passé, *via* des séries de data bien réelles. Si le développement des méthodes d'auto-apprentissage, de type *machine learning*, vient progressivement bouleverser le paysage, il n'en reste pas moins qu'il faut partir du passé pour penser le futur. Un algorithme, aussi sophistiqué soit-il, reste prisonnier du passé. Dès qu'un évènement nouveau ou extrême arrive, qui n'a pas été modélisé, éveillez votre attention. Il y a de très fortes probabilités que la machine ne suive plus. Il faudra du temps pour qu'elle se remette de ses émotions.

QUEL AFFREUX JARGON !

MACHINE LEARNING

Désigne des algorithmes qui adaptent, de façon continue, leurs comportements ou leurs réponses, en se fondant sur des data empiriques, sans avoir été nécessairement programmés *a priori*.

Prenez un exemple simple : votre objectif est de concevoir un modèle qui permette de convertir, automatiquement, des degrés Celsius (en entrée) en degrés Fahrenheit (en sortie). Deux options s'offrent à vous : la première option, c'est l'approche classique. Elle vous amènera à coder en dur (et *a priori*) la formule de conversion, en l'occurrence : F = (C x 1,8) + 32, puis à passer chaque degré Celsius que vous souhaitez convertir à la moulinette de cette formule. Là où le bât blesse, c'est que vous ne

■■■

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

connaissez pas, dans la plupart des cas, les formules *a priori*. Dans ces cas-là, c'est la seconde option, celle du *machine learning*, qui s'imposera. Imaginez que vous ne connaissiez pas la formule F = (C x 1,8) + 32. Dans la démarche *machine learning*, vous vous contenterez de nourrir le modèle de quelques valeurs de correspondance Celsius / Fahrenheit déjà observées… et le modèle en déduira, à partir de ces quelques exemples, une approximation de la formule idoine, toujours plus précise. Sans avoir à coder la formule *a priori*, la machine la déduit *a posteriori*. Elle l'appliquera aux prochaines valeurs en Celsius que vous lui apporterez.

A été évoquée, dans la partie « La data, pour ne pas disparaître ! » (p. 23-30), l'importance, pour chacun d'entre nous, de comprendre les algorithmes qui nous entourent. Pas dans le détail, mais dans ses grandes lignes : *Quel est son objectif final ?* Cette exigence ne doit pas rester uniquement celle des citoyens, mais devenir celle de tous vos collaborateurs en contact avec de la data. Disons-le franchement : cette exigence de compréhension sera toujours plus difficile à tenir, dans un contexte de développement massif des algorithmes de type *machine learning*. Ces modèles gagnent en précision ce qu'ils perdent en intelligibilité. Mais pour votre, le jeu en vaut la chandelle !

UN PAS DE CÔTÉ !

Pour créer cette vigilance collective sur les data et les algorithmes, l'Éducation Nationale a un rôle à jouer. Nous lui suggérons, dès que possible, d'intégrer aux programmes scolaire l'apprentissage du code informatique et de la statistique.

Apprendre aux enfants à coder, d'abord, pour soulever le voile, et comprendre ce qui se cache derrière les écrans. Les humanités éclairent sur la vérité des cœurs, les sciences économiques sur les composantes de la croissance du PIB. Quelles sciences pour démys

Copyright © 2018 Dunod.

tifier le digital ? Comprendre le code, car sinon le monde restera complexe, obscur et laissera place aux plus farfelues superstitions. Pour reprendre et tordre les mots de Baudelaire, le monde digital est *Un temple où de vivants piliers / Laissent parfois sortir de confuses paroles / L'homme y passe à travers des forêts de symboles*. Nos enfants doivent comprendre le monde qui les entoure, pour ne pas rester ahuris, crédules, devant les magiciens. Il faudrait donc enseigner le code comme une science dure, non connectée.

Apprendre à scorer (faire de la statistique), enfin, et enseigner la statistique, pour comprendre nos propres décisions. Nos actions quotidiennes, en ligne, sont déjà dictées par des algorithmes : le livre recommandé sur Amazon ; la publicité en premier résultat sur Google ; jusqu'à, très bientôt, nos conduites en voiture. Ces micro-ajustements, demain, viendront assécher nos libertés individuelles. Vous souhaitez devenir végétarien ? Les entreprises d'agro-alimentaires pourront, *via* des algorithmes publicitaires sophistiqués, subrepticement vous ramener dans le droit chemin, en vous suggérant une promotion anodine pour la viande bovine lorsqu'ils vous trouveront sur Google. Nos enfants doivent savoir comment fonctionnent les algorithmes, et dans quelle mesure ils nous entourent et influencent nos actions.

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

Copyright © 2018 Dunod.

Conclusion

Il vous faut bouger fort (et vite !) sur la data. C'est notre conviction. Nous espérons qu'elle sera contagieuse. La data n'est pas un nouveau « buzzword » à la mode, comme la *blockchain*[1], les *chatbots*[2], ou autres têtes d'affiche des rapports Gartner. C'est une lame de fond, comme le digital. Digital et data ? Ce sont d'ailleurs les deux faces d'une même médaille, d'une même révolution. Les deux constituent le nouvel environnement, demain omniprésent, de votre activité. Digitaliser, c'est produire des data. Automatiser, c'est structurer, en séquence, des data. Concevoir de l'intelligence artificielle, c'est, *in fine*, laisser jouer des data, entre elles, en autonomie.

À vous de jouer !

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

1. Base de data distribuée transparente, sécurisée, et fonctionnant sans organe central de contrôle.
2. Robot logiciel pouvant dialoguer avec un individu ou consommateur par le biais d'un service de conversations automatisées.

Copyright © 2018 Dunod.

Bibliographie

Babinet Gilles, *Big Data, penser l'homme et le monde autrement*, Le Passeur, 2015.

Cardon Dominique, *À quoi rêvent les algorithmes*, La République des Idées, 2015.

Carr Nicholas, *The Shallows*, Atlantic Books, 2010.

Carr Nicholas, *The Glass Cage*, Penguin, 2016.

Cointot Jean-Charles et Eychenne Yves, *La Révolution Big Data – Les données au cœur de la transformation de l'entreprise*, Dunod, 2014.

Morozov Evgeny, *To Save Everything, Click Here: The Folly of Technological Solutionism*, Penguin, 2013.

Morozov Evgeny, *Le Mirage numérique : pour une politique du Big Data*, Les Prairies Ordinaires, 2015.

Newport Cal, *Deep Work: Rules for Focused Success in a Distracted World*, Piatkus, 2016.

Sadin Eric, *La Vie algorithmique : Critique de la raison numérique*, 2015.

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.

Copyright © 2018 Dunod.

Webographie

https://france.emc.com/campaign/observatoire-de-la-donnee/index.htm

www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

www.alliancy.fr/a-laffiche/relation-client-crm/2016/04/27/carrefour-media-connecte-les-donnees-pour-les-marques

www.alliancy.fr/a-laffiche/relation-client-crm/2016/04/27/carrefour-media-connecte-les-donnees-pour-les-marques

www.ladn.eu/entreprises-innovantes/transparence/publicite-en-ligne-amazon-sort-son-epingle-du-jeu

www.lesechos.fr/tech-medias/medias/030427306723-pub-en-ligne-les-medias-francais-font-front-commun-2099718.php

http://ideas.microsoft.fr/cybersecurite-5-chiffres-cles-a-connaitre/#eHvxw3 UvPvbG4IqR.97

www.kaushik.net/avinash/kill-useless-web-metrics-apply-so-what-test

www.economist.com/news/science-and-technology/21694992-scientific-data-might-be-filled-important-things-waiting-be-discovered

https://techcrunch.com/2016/02/01/gmail-now-has-more-than-1b-monthly-active-users

https://medium.com/airbnb-engineering/superset-scaling-data-access-and-visual-insights-at-airbnb-3ce3e9b88a7f

Copyright © 2018 Dunod.
© Dunod - Toute reproduction non autorisée est un délit.